POWER PREDICTION USING THE WIND TURBINE POWER CURVE AND DATA-DRIVEN APPROACHES

EHSAN TASLIMI RENANI

INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

POWER PREDICTION USING THE WIND TURBINE POWER CURVE AND DATA-DRIVEN APPROACHES

EHSAN TASLIMI RENANI

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Ehsan Taslimi Renani

Matric No: HHD120007

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Thesis ("this Work"): POWER PREDICTION USING THE WIND TURBINE POWER CURVE AND DATA-DRIVEN APPROACHES

Field of Study: Energy

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

POWER PREDICTION USING THE WIND TURBINE POWER CURVE AND DATA-DRIVEN APPROACHES

ABSTRACT

Wind energy as one of the promising energy sources, has attracted great attention because it is pollution-free and abundant. Moreover, it shows considerable potential for supplying electricity to meet the demand. The high dependence upon the wind, however, results in variation of the wind power due to the intermittent nature of the wind. The volatility of wind power over time jeopardizes the reliability of the power systems. Therefore, the prediction of the wind power is required. Wind turbine power curve representing the relationship between the wind speed and power can serve as a tool for prediction. In this thesis, a new parametric model, called modified hyperbolic tangent (MHTan), is proposed to approximate the wind turbine power curve. To obtain the unknown vector of parameters of the MHTan, three heuristic optimization algorithms are employed to minimize the sum of squared residuals. An alternative way to estimate the coefficients of MHTan is through maximum likelihood estimation (MLE) and the probability density function of wind speed. In this method, firstly, Weibull density function is utilized to model the wind speed and then several methods are applied to estimate the parameters of the wind speed distribution. To evaluate the performance of the Weibull parameters' estimator methods, two sets of data are considered, one based on simulated data with different random variable size and the other based on actual data collected from a wind farm in Iran. Secondly, a new formula representing frequency distribution of the turbine power is derived. The formula comprises of unknown vector parameters of MHTan which can be determined based on MLE. Then, the performance of the MHTan is evaluated using actual data collected as well as three simulated data representing three different typical shapes of the power curve. In order to demonstrate the efficiency of the proposed method, it is compared rigorously with several parametric and nonparametric models. In addition, the capability of the MHTan in on-line monitoring of the wind turbine is presented. In this research, a comparison is also drawn between two different wind power prediction models, indirect and direct approaches. In the former it is necessary to forecast the wind speed at first, then the corresponding power is obtained from the wind-power curve. Since in practice turbines do not work in ideal conditions, the theoretical power curve provided by manufacturers is avoided and a power curve approximated by MHTan is used instead. Several statistical methods are used to predict wind speed and the best one is selected for prediction over longer horizons. To set up direct wind power prediction, six data-driven approaches are employed and the same procedure as in indirect approach is applied to select the best method for longer horizon predictions, up to 60-min. The results confirm the superiority of the direct prediction models. Moreover, a hybrid feature selection technique is proposed to choose the necessary subset of inputs so that the important information is retained. This technique is a combination of mutual information and neural network where its effectiveness is examined with several linear and nonlinear feature selection methods.

Keywords: wind power prediction, wind turbine power curve, wind speed prediction, data-driven

POWER PREDICTION USING THE WIND TURBINE POWER CURVE AND DATA-DRIVEN APPROACHES

ABSTRAK

Tenaga yang boleh diperbaharui, terutamanya tenaga angin, telah banyak meningkat sejak sedekad yang lalu, disebabkan oleh pencemaran alam sekitar dan pengurangan sumber bahan api fosil. Tenaga angin sangat penting kerana potensi yang besar dalam membekalkan tenaga elektrik untuk memenuhi permintaan. Walau bagaimanapun, pergantungan tinggi pada angin menimbulkan kecenderungan perubahan kuasa angin melalui sifat rawak dan stokastik tenaga angin. Tambahan pula, kerapuhan dan ketidakstabilan kuasa angin dari masa ke masa mengurangkan kebolehpercayaan sistem kuasa. Oleh itu, ramalan kuasa angin adalah suatu keperluan. Keluk kuasa turbin angin yang mewakili hubungan antara kelajuan angin dan kuasa boleh berfungsi sebagai alat untuk ramalan. Dalam tesis ini, model parametrik baru, yang dikenali sebagai tangen hiperbolik diubah suai (MHTan), dicadangkan untuk menghampiri keluk kuasa turbin angin. Untuk mendapatkan vektor parameter yang tidak diketahui MHTan, tiga algoritma pengoptimuman heuristik digunakan untuk meminimumkan jumlah sisa kuasa dua. Cara alternatif untuk menganggarkan koefisien MHTan adalah melalui estimasi kemungkinan maksimum (MLE) dan fungsi kepadatan kebarangkalian kelajuan angin. Dalam kaedah ini, beberapa kaedah pertama digunakan untuk menganggar parameter pengagihan kelajuan angin. Di sini fungsi Weibull digunakan untuk memodelkan kelajuan angin, walau bagaimanapun fungsi ketumpatan lain juga diperiksa. Untuk menilai prestasi kaedah penganggar parameter Weibull, dua set data dipertimbangkan, satu berdasarkan data simulasi dengan saiz variabel rawak yang berbeza dan yang lain berdasarkan data sebenar yang dikumpulkan dari ladang angin di Iran. Kedua, formula baru yang mewakili pengagihan frekuensi kuasa turbin diperolehi supaya parameter yang tidak diketahui dalam formula ini memang merupakan pekali MHTan yang boleh

diperolehi dengan terbitan MLE. Prestasi MHTan kemudiannya dinilai menggunakan data sebenar yang dikumpulkan serta tiga data simulasi yang mewakili tiga bentuk tipikal kurva kuasa. Untuk menunjukkan kecekapan kaedah yang dicadangkan, ia dibandingkan dengan beberapa model parametrik dan bukan parametrik. Di samping itu, kapasiti MHTan dalam pemantauan dalam talian turbin angin dibentangkan. Dalam kajian ini dibuat juga perbandingan antara dua model ramalan kuasa angin yang berbeza, pendekatan tidak langsung dan langsung. Di dalam pendekatan tidak langsung, ramalan kelajuan angin adalah perlu pada mulanya, di mana kuasa sepadan boleh diperoleh melalui lengkung daya angin. Oleh kerana turbin tidak berfungsi dalam keadaan yang ideal, lengkung uasa teori yang disediakan oleh pengeluar dihindari dan keluk kuasa yang dihitung oleh MHTan digunakan sebaliknya. Terdapat beberapa kaedah statistik berfungsi untuk meramalkan kelajuan angin dan yang terbaik dipilih untuk ramalan bagi jangka masa yang lebih lama. Untuk menetapkan ramalan kuasa angin secara langsung, enam pendekatan didorong data digunakan dan prosedur yang sama seperti pendekatan tidak langsung digunakan untuk memilih kaedah terbaik untuk ramalan bagi jangka masa yang lebih panjang, sehingga 60-minit. Keputusan mengesahkan keunggulan model ramalan langsung. Selain itu, teknik pemilihan ciri hibrid dicadangkan supaya memilih subset input yang diperlukan supaya maklumat penting masih dikekalkan. Teknik ini adalah gabungan maklumat bersama dan rangkaian saraf dan keberkesanannya diperiksa dengan beberapa kaedah pemilihan ciri linier dan tak linear.

Kata kunci: ramalan kuasa angina, lengkung kuasa turbin angina, ramalan kelajuan angina, didorong data

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to Prof. Ir. Dr. Nasrudin Bin Abd Rahim and Dr. Mohamad Fathi Mohamad Elias for their supervision, guidance, continuous encouragement, helpful support, financial support, and valuable comments throughout the progress of my research and the writing of this thesis. I would like to extend my appreciation to UM Power Energy Dedicated Advanced Centre (UMPEDAC) and Institute of Graduate Studies, University of Malaya, for the rich research resources.

Furthermore, I am very grateful to my family for their patient and continuous encouragement that has enabled me to complete this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Pop	:	Population matrix
U	:	Uniform distribution
histPop	:	Historical population matrix
F	:	Amplitude control function of search-direction matrix
TR	:	Final population matrix
Mutant	:	Trial population matrix
map	:	Binary matrix
nPop	:	Population size
D	:	Dimension of the search space
mixrate	:	BSA's control parameter
$Gbest^t$	:	Global best position at iteration t
$Pbest_j^t$	:	The best position of individual j at iteration t
$\boldsymbol{V}_j$	:	Speed vector of jth particle
ρ_1, ρ_2	:	Random numbers uniformly distributed
c_1, c_2	:	Setting parameters of PSO
w	:	Inertia weight
ϑ	:	CSA's step size parameter
$rand_x$, $rand_y$	:	Normally distributed stochastic variables
Γ	:	Gamma distribution function
κ	:	Distribution factor in CSA
c, k	:	Parameters of Weibull distribution
σ	:	Standard deviation
μ	:	Mean value
$\boldsymbol{\varphi}_i^t$	:	Head angle of ith member at iteration t

$\mathbf{D}_i^t$	:	Search direction of ith member at iteration t
$\beta_{\max}$	:	The maximum pursuit angle in CSA
$l_{\max}$	:	The maximum pursuit distance in CSA
θ	:	Unknown parameters of parametric models of wind turbine power curve
n	:	Sample size
y_a	:	Actual value
y_e	:	Estimated value
α	:	Significance level
ϱ	:	Control parameter in on-line monitoring
η, λ	:	Smoothing parameters in DES
P	:	Order of the auto-regression
q	:	Order of the moving average
ψ	:	Sigmoid activation function
ζ	:	Tangent hyperbolic transfer function
V	:	Number of selected attribute
w_l^m	:	Connection weight from l-th attribute to m-th hidden unit
v_o^m	:	Connected weight from m-th hidden unit to the network output
$\alpha_1, \alpha_2, \varsigma$	:	Control parameters of penalty term
ξ, ξ_i^*	:	Slack variables in SVM
C	:	Regularization parameter in SVM
ϵ	:	Tolerance threshold in SVM
α_i^*, α_j	:	Lagrange multipliers
γ	:	Membership function
χ^2	:	Chi-square test

p.e	:	Percentage error
df	:	Degree of freedom
$H(X)$	:	Entropy of random variable X
$P(X)$	:	Probability function of X
$H(X,Y)$	:	Joint entropy of X and Y
$P(X,Y)$	:	Joint probability distribution of variables X and Y
$H(X\|Y)$	:	Conditional entropy
$I(X;Y)$	:	Mutual information of variables X and Y
K-S	:	Kolmogorov–Smirnov test
R^2	:	Coefficient of determination
T	:	Temperature
5-PL	:	Logistic five-parameter
4-PL	:	Logistic four-parameter
ACF	:	Auto-correlation function
ANFIS	:	Adaptive neuro-fuzzy inference system
ARMA	:	Auto-regressive moving average
BSA	:	Backtracking search algorithm
CA	:	Correlation analysis
CDF	:	Cumulative density function
CSA	:	Cuckoo search algorithm
DES	:	Double exponential smoothing
EM	:	Empirical method
EPFM	:	Energy pattern factor method
Gd	:	Grid partition
GM	:	Graphical method
GSO	:	Group search optimization

IS	:	Information sharing
k-NN	:	k-nearest neighbor algorithm
LSE	:	Least squared error
MAE	:	Mean absolute error
MAPE	:	Mean absolute percentage error
MHTan	:	Modified hyperbolic tangent
MI	:	Mutual information
MLE	:	Maximum likelihood estimation
MLP	:	Multilayer perceptron
MM	:	The method of moment estimation
NN	:	Neural network
NWP	:	Numerical weather prediction
PCA	:	Principal component analysis
PDF	:	Probability density function
PSO	:	Particle swarm optimization
RBF	:	Radial basis function
RF	:	Random forest
RMSE	:	Root mean square error
SDE	:	Standard deviation error
SVM	:	Support vector machine
UCL	:	Upper control limit
LCL	:	Lower control limit
WD	:	Wind direction
WP	:	Wind power
WS	:	Wind speed
WTPC	:	Wind turbine power curve

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Introduction

Wind is one of the fastest growing energy sources because it is renewable and pollution-free. Other factors such as advances in manufacturing and control technology intensify appeal of the wind as a green source of energy. According to Global Wind Energy Council (GWEC), the accumulative total of wind energy installed capacity reached approximately 486 GW in 2016, representing a growth rate of 12.5% than that in the previous year (Council, 2016). European Wind Energy Association expects 392 GW installed capacity by 2030, equal to 31% of European electricity demand and avoidance of 554 Mt of CO_2 emission caused by conventional electricity generation. High penetration of wind power generation, however, provides great challenges in electrical power systems due to stochastic nature of the wind flow. Although in the power electricity grid, supply must meet power demand all times, the fluctuation of wind power output makes difficulties to maintain this balance. The variation of wind energy not only jeopardizes quality and stability of the power systems, but also affects the wind energy providers who are imposed penalty because of the failure of generating the contracted amount of energy. Moreover, it causes significant uncertainties to transmission system operators (TSOs) who need precise information for unit commitment.

The increased wind energy sources call for wind power prediction to alleviate the undesirable effects of the wind energy integration into electric power grids. Wind power forecasting aims to help TSOs to effectively determine the reserve power in order to balance possible errors between programmed and actually generated wind power in a certain time period.

1.2 Problem Statement

Quality of forecasts is very important and thus improving prediction systems' performance has been set as one of the priorities in wind energy research. Persistent method, known also as 'Naïve Predictor', is the simplest model in wind power prediction. However, it only illustrates accurate results in very short term forecasting. Some researches attempt to predict wind power through wind turbine power curve (El-Fouly, El-Saadany, & Salama, 2007). In this regard, firstly, wind speed at wind turbines locations must be forecasted then it is converted into power through the characteristic curve of the wind turbine. Wind power curve either can serve as a tool for predicting wind power or aids in performance monitoring of the turbine. Indeed, any anomalies can be effectively detected by monitoring wind turbines. Wind turbine manufacturers although provide power curve, there is a substantial discrepancy between the theoretical and the empirical power curves because the wear and tear of turbine generator such as gearbox as well as local turbulence are disregarded. Moreover, wind speed distribution, location of the turbine and wind direction may exacerbate this difference. In practice, in fact, wind turbines rarely, if ever, operate in ideal conditions.

Models for predicting the behavior of the turbine include: models based on fundamental equations of power available in the wind and models based on the concept of the power curve of wind turbine. Different models based on the equation of power were presented, but they are highly dependent on various parameters such as wind speed, turbine rotational speed, turbine blade parameters (pitch angle and angle of attack), efficiency of generator and mechanical transmission (Ashok, 2007; Nelson, Nehrir, & Wang, 2006; Hongxing Yang, Wei, & Chengzhi, 2009). Due to the unavoidable dependency of these parameters and their variation with change in weather condition and type of component used, these models do not yield accurate results and are also cumbersome. Models based on the concept of the power curve of turbine comprise of parametric and nonparametric approaches. Polynomial (Chedid, Akiki, & Rahman, 1998), logistic four-parameter (Kusiak, Zheng, & Song, 2009b), logistic five-parameter (Lydia, Selvakumar, Kumar, & Kumar, 2013) are typical examples of parametric approaches, which use the mathematical expression to estimate wind power curve. Nonparametric approaches such as copula power curve (Gill, Stephen, & Galloway, 2012), cubic spline (Shokrzadeh, Jozani, & Bibeau, 2014), data mining (Schlechtingen, Santos, & Achiche, 2013), neural network (G. Sideratos & N. D. Hatziargyriou, 2007) and fuzzy methods (M. Mohandes, Rehman, & Rahman, 2011) attempt to find a relationship between wind power and wind speed. Such methods have a major disadvantage over parametric approaches as they suffer from the black box problem. Indeed, their output is difficult to backtrack. Moreover, some of them as NNs fall into local extremum due to the overfitting problem. Hence, producing a new wind turbine power curve model which is as close as possible to the observed data is greatly encouraged.

Several studies were developed to predict wind power without involvement in the wind turbine power curve. They employed historical wind speed, wind direction, humidity, and also previous power generated as inputs of the forecast engine to predict next step wind power. Usually in these researches, selection of important inputs was disregarded. Selection of most important inputs improves the results. Furthermore, in case of engine forecast is kind of machine learning, it speeds up the learning process and avoids overfitting the training data. (Fan, Liao, Yokoyama, Chen, & Lee, 2009) applied cross correlation and auto correlation to select exogenous variables having a stronger relationship with the wind generation. However, it was not able to evaluate correctly the impact of each input variable on wind power. It is due to the fact that cross correlation and autocorrelation are linear feature selections whereas wind power is a nonlinear mapping function of its input variables. Indeed, using a nonlinear analysis technique for feature selection and the combination of it with the forecast engine is a great need to improve prediction accuracy.

1.3 Objectives

The main objectives of this study are:

- To propose a parametric model to approximate the wind turbine power curve in order to serve as prediction tool in indirect wind power prediction and wind turbine monitoring
- To optimize the unknown coefficients of the proposed parametric model
- To estimate the parameters of the proposed parametric model through wind speed distribution
- To validate the proposed parametric wind turbine power curve model for vertical and horizontal axis wind turbines with various power-curve shapes
- To propose a two-stage feature selection technique as a preprocessing tool in direct wind power prediction using combination of the mutual information and neural network to extract the most informative features with a maximum relevancy and minimum redundancy

1.4 Scope of Work

The following items are considered in this study:

- Estimation of the wind turbine power curve using both parametric and nonparametric approaches is considered.
- Performance evaluation of the parametric and nonparametric based on the theoretical power curve data supplied by the manufacturer and also data observed from a real wind farm.

- Direct and indirect techniques are investigated to forecast the wind power generation.
- In indirect prediction approach, three statistical methods are considered to forecast the wind speed and to set up the direct prediction approach six data-driven approaches are applied.

- In direct approaches, the feature selection technique is employed to filter out irrelevant and redundant inputs in the procedure of power estimation. The feature selection technique is applied on real meteorological data over a period of one year, including wind speed, temperature, wind power generated, and wind direction.
- The simulations and analysis are carried out in MATLAB and WEKA on a personal computer with Intel Pentium 2.66 GHz processor and 4GB RAM.

1.5 Organization of Thesis

This thesis is divided into five chapters as follows:

Chapter 1 contains problem statement, research objectives, the scope of work, and an overview of the thesis.

Chapter 2 presents a literature survey of the wind power forecasting methods as well as wind turbine power curve modeling.

Chapter 3 focuses on developing a model for wind turbine power curve. It investigates two different approaches to estimate the parameters of the wind turbine models. This chapter also derives a statistical model to match the frequency distribution of the wind power. Then it describes the comparative results.

Chapter 4 compares direct and indirect wind power prediction approaches. An intelligent feature selection technique is presented in this chapter in order to select the most informative inputs. It also describes the results of wind speed and wind power prediction. Then, comparative analysis of the power predicting models and feature selection technique is discussed.

Chapter 5 presents the thesis conclusion and suggests the future work. A comprehensive list of reference is provided at the end of the thesis.

CHAPTER 2: LITERATURE REVIEW

2.1 History of Wind Power

The usage of the wind as a source of power has been dated about 5000 B.C. when the first sailing boats were used on the Nile River. The Persian Empire, too, had utilized wind energy to pump water and grind grain. In 1887, the first windmill was constructed by Prof. James Blyth, in Scotland to produce electricity for his personal cottage. Shortly thereafter, Charles F. Brush completed the construction of the first fully automatic windmill having the power of 12 kW. The mass production of wind machines, from 5 to 25 kW in size, began in America by the end of the 1900s. In the early of 20 centuries, the term “wind turbine” greatly increased popularity for describing a machine used to convert kinetic energy in the wind into electrical power. In this era, an enormous amount of wind turbine, 600,000 units approximately, were installed in U.S. Primary turbines were explored showing different design alternatives including vertical axis turbines (VAWTs). As the industry matured, horizontal axis wind turbines (HAWTs) were standardized with one, two, or three blades. In 1941, a modern horizontal axis wind turbine was installed in U.S to provide electricity to a remote area where the electric power lines could not reach. Despite the development of electricity power lines and the technological advancement in megawatt turbines, wind turbines market fell into a decline in the 1950s. Due to the oil crisis and its skyrocketed price, wind turbines came to the sharp focus again in the 1970s. Since the beginning of the 21st century, wind experienced a great leap in usage. According to world wind energy association, by the end of 2016, more than 85 countries in the world are using wind power on a commercial basis. The global wind installed capacity from 2000 to 2016 is illustrated in Figure 2.1 (Association, 2014). It is shown that, at the end of 2016, the global wind energy installed capacity dramatically enhanced to 486 GW, a 300% growth over 2008.

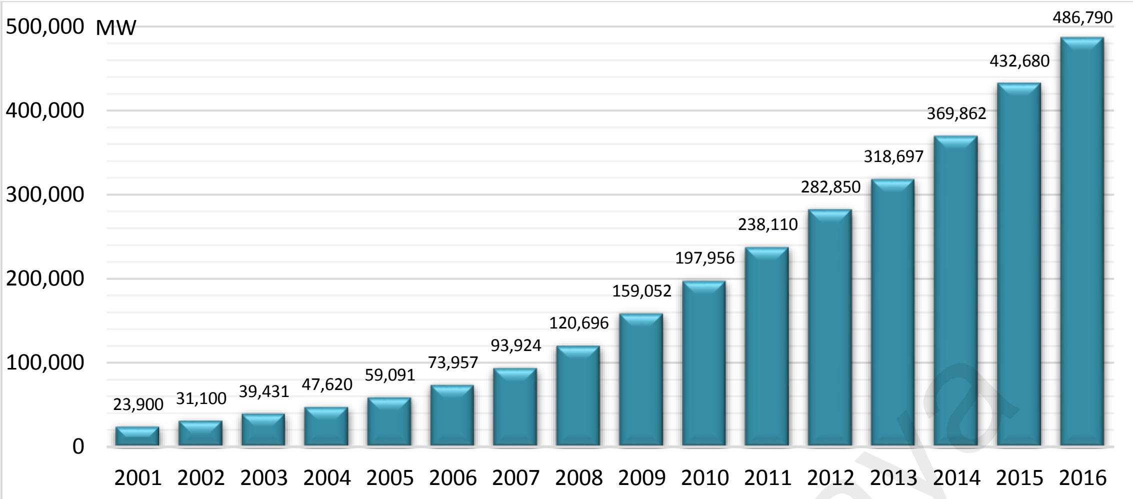


Figure 2.1: Global cumulative installed wind capacity 2000-2016

2.2 The Importance of Wind Power Forecasting

Driven by depletion of fossil fuel resources, energy independence, and contamination of the environment, renewable energy most notably wind energy is becoming increasingly important as one the most promising alternative source of energy. The primary motivation behind the widespread usage of wind energy is to reduce green gas emission which is not only considered a growing menace to the Earth but also to human health. Electricity generation produces the largest share of greenhouse gas emissions, about 30 percent of emitted greenhouse gas, due to the burning fossil fuels, mostly oil, coal, and natural gas.

Wind energy is a green and polluted-free resource and it becomes more and more cost-effective by virtue of the technologies development. Since this energy resource depends on the wind, it has no fuel cost. In other words, the motion of the air rotates the wind turbines, converting kinetic energy into mechanical energy and then electricity is converted from mechanical energy by the rotating magnetic coils in the gearbox. In this procedure, the electricity is produced without emitting any greenhouse gas and the only input needed is the wind which is free and rich in nature. Moreover, wind turbines and towers can be recycled without making any pollution. Recent innovation in wind industry enables us to capture stronger wind in higher altitude and offshore, and consequently greater amount of power can be generated by the turbine. Despite all the advantages of wind energy, there are diverse technical and economic issues in the integration of wind power that must be addressed. One issue is that unlike conventional hydro-thermal generation, wind industry mainly uses the asynchronous machine which requires power electronic as an interface between the grid and the generator. The most severe issue in the wind industry, however, is deeply rooted in the random character of the wind energy which imposes great difficulties to manage electricity power grid. In other words, high dependency on the wind gives rise to the mutability of the wind power through the stochastic nature of the wind.

The inherent intermittency and variability of wind generate uncertainty about the real production of a wind farm. Wind power, unlike conventional power plant, is not fully dispatchable. A fossil-fuel plant, for instance, can adjust its output to the demand or even can be turned off and on, while in wind power plants the output cannot be controlled in command of power system operators. Moreover, the wind cannot be stored as other resources such as coal and natural gas for future power generation. Wind energy is considered as a fluctuating source of electrical energy which makes difficulties for transmission operators (TSOs) as well as wind power plant (WPP) owners who need robust information for dispatching, unit commitment and decision-making in the electricity market. TSOs have an obligation to ensure that generation and consumption levels remain adequate at all times. In fact, they must guarantee a good balance between demand and power supply in an unbiased manner at the optimum cost under the constraints of the transmission network. Mutability of the wind power on one side and the balance between load and supply on the other side obliges TSO to find a solution. Although one alternative to mitigate this issue is to use energy storage, it appears not to be feasible for large-scale wind generation. The most common solution is to assign reserve capacity (spinning reserve) to balance demand and power generation. The spinning reserves (SR) allow power system operators to compensate any unpredictable unbalanced between consumption and generation. For instance, if wind power turbines fail to produce electricity as much as expected due to the variability of the wind speed other fast responding units, which are mostly gas-fired power plant, are needed to balance the network. The cost of SR, however, is far from negligible. In addition, the higher integration of the wind into the power grid, the more the calls for conventional power resources to cover the gap between load and demand. In real-time operation, looking seconds to hours ahead, in order to balance this variability traditional hydro-thermal generations need to ramp up or down. A frequent ramping action not only has adverse economic effects such as mechanical wear and tear but also reduces the efficiency in the use of conventional hydro-thermal generation. Another controversial point regarding integration of wind power is the location of wind farms. Wind parks should be only located in areas with good wind regimes, however, these are sometimes remote area and far from existing transmission infrastructure hence, the grid improvements turn out to be very expensive.

2.3 Factors Affecting Wind Power Production

The theoretical power captured by the rotor of a turbine is given by:

$$P = 0.5\rho\pi R^2 C_p(\lambda,\beta)V^3 \qquad (2.1)$$

$$C_p(\lambda,\beta) = 0.5176\left(\frac{116}{\frac{1}{\lambda-0.08\beta}-\frac{0.035}{\beta^3+1}} - 0.4\beta - 5\right)e^{\frac{-21}{\frac{1}{\lambda-0.08\beta}-\frac{0.035}{\beta^3+1}}+0.0068\lambda} \qquad (2.2)$$

where ρ is the air density kg/m^3, R is the radius of the rotor in m determining its swept area, C_p is the power coefficient and V^3 is the wind speed in m/s. Representing the percentage of the power captured by the turbine is function of the turbine tip speed ratio λ and the blade pitch angle β. Smaller pitch angles result in higher power coefficients and thus higher energy output. According to the Betz's law, only 59.3% of the kinetic energy in wind can be captured by wind turbines. ($C_p \leq 0.593$).

2.3.1 Air Density

The wind power is directly proportional to air density and any change in the air density affects annual energy output. The air density is strongly related to the pressure, the humidity and the temperature. The air is denser at lower elevations and cold air is denser than warm air. The lower air density is, the weaker strength of wind will be, and then the starting and rated wind speed will be increased, and consequently output power will be less. On the contrary, the output energy will be increased when the air density is higher.

2.3.2 Wind Speed

According to Eq. (2.1), the output power of the wind turbine is proportional to the cube of the wind speed. Therefore, small changes in wind speed make the significant differences in power. Wind speed increases with height hence if the tower of the wind turbine is taller, the wind speed and in turn, energy output of the turbine will be higher. In general, the most crucial data required to calculate the output energy of the wind generator is the wind speed at the particular site. A rapid change in wind speed called wind shear which may be horizontal or vertical. Vertical wind shear is the rate of change of the wind speed with a change in altitude, while horizontal wind shear is the rate of change of the wind speed in the horizontal plane. Both have an adverse effect on the output power generated by the turbine. Vertical wind shear, for example, causes different wind speeds passing through the two blades, nearest and farthest to the ground level and this in turn strongly affects the wind turbine operation.

2.3.3 Temperature

The temperature and pressure are vital factors influencing the wind power since both parameters affect the air density. Ice might be accumulated on the wind turbine rotor disc, if the temperature is too cold, resulting in a failure in turbine operation. Low temperature, moreover, can damage the electrical equipment of the turbine as well as affecting the lubricant which in turn reducing power transmission of the gearbox.

2.3.4 Icing

Icing denotes a serious threat to the integrity of wind turbines in cold weather. Lift reduces and drags increases along the wind turbine blade following the power law. Under severe icing condition, accumulated ice on the wind turbine blades reduces the torque and aerodynamic efficiency resulting in power loss. In icing weather, torque may dramatically drop to almost zero and the wind turbine fails to operate normally and thus an utter loss of power production. In general, two types of icing which are likely to form on the wind turbine blades: glaze and rime. The former occurs when liquid precipitations striking the surface at a temperature lower than freezing point. Glaze ice is a common occurrence during ice storms and is relatively transparent, smooth, hard, and attaches well to surfaces. Rime ice is formed when a supercooled droplets freeze on contact with a surface below freezing point. Since the droplets are small, they instantly freeze and create a mixture of trapped air and tiny ice particles. This type of icing is rough crystalline structure, and brittle. Rime ice can cover airfoil surface and consequently affect the aerodynamic characteristic of the blades. It is of vital importance of in high elevation location, e.g., hills or mountain tops.

2.3.5 Dust

Blade surface roughness of the wind turbine has a significant impact on the aerodynamic loads and wind power generation. The increase in the drag force of the airfoil, as well as the decrease in the lift force as a consequence of growing dust on the surface of the turbine blades, diminish the energy output of the wind turbine. In other words, gathered dust on the turbine blades degrades the smoothness of its surface and hampers the airfoil to extract the useful power from the wind result in a greater loss of the wind turbine output power.

2.3.6 Wake Effect

Considering wind turbines extract the kinetic energy in wind and convert it into the electricity, the wind leaving the turbine contains less energy than the wind upstream of the turbine. Consequently, the wind downstream of a wind turbine has lower speed and higher turbulence as compared to the wind in the free stream. This downstream wind is the wake of the turbine. This turbulent and slow downed wind from an upwind turbine reduces energy arriving downwind turbines and, in consequence, the overall output energy of the downwind turbine decreases. Although increasing the space between turbines decreases the wake effect on downstream wind turbines, it is limited by land and excessive cost of cabling. Two significant influences of wake are: (i) diminution of wind velocity which reduces, in turn, the energy generation of the wind farm; (ii) a rise in the turbulence of the wind which increases mechanical loads on downwind turbines and diminishes their operational capacity. In practice, however, the separation between turbines of 3 or 4 diameters apart is specified to counteract the wake effect, but blades characteristics as well as environmental factors such as humidity, temperature, complex terrain, and forestry may affect the magnitude and size of wakes.

2.3.7 Direction

Although the wind direction has been disregarded in many types of research, it has a substantial influence on wind farm performance. Since wind farm layout and wind direction can modify the location and the orientation of the wake cones, any changes in these factors can affect wake interaction (overlapping area) which leads to changes in the power output of each individual turbine. (Pinson, 2006) illustrated that variation of the wind direction has a more considerable influence on the wake coefficient in wind farms with a compact arrangement of the wind turbines than in wind farms with larger wind turbines spacing.

2.4 Wind Power and Wind Speed Forecasting Techniques

Stochastic nature of wind power generation poses great challenges on wind power systems. For instance, in power systems with a high share of wind power, the intermittency of the wind could oblige power system operator to allocate greater supplemental energy reserve to keep a balance between load and generation and minimize the errors between programmed and actual generated energy by wind power in a certain period of time. This would, however, impose an additional operation cost and consequently increase the final energy price. On one hand, the variation of wind energy can jeopardize the quality and stability of power systems and on the other hand, it affects the market participants who bear the economic losses because of the failure of generating the contracted amount of energy. Hence, an accurate wind forecasting model as an efficient tool to save costs and effectively support Distribution and Transmission System Operators (DSO/TSO) in improving power network management is required.

In the recent years, a wide range of studies has been carried out on wind power forecasting, each using different techniques with different prediction time horizon. There are several ways to categorize the wind power forecasting approaches. Recent researches in the area of wind power prediction mainly are classified into three groups, statistical, physical and hybrid approaches. The first methods attempt to tune model parameters to minimize the error between the observed and the predicted power based on the vast historical data. Physical approaches, however, depend on numerical weather prediction and meteorology information i.e., obstacles, temperature, terrain, and pressure. The last group incorporates the individual superiority of the diverse prediction models so as to improve the accuracy of the forecasting.

Forecasting models for wind power, however, can be also categorized as direct and indirect models. The first approach attempts to forecast wind power generation directly from previously recorded data consisting of temperature, humidity, wind speed, wind power, and wind direction, . The latter is based on predicting wind speed at turbine location then using the wind turbine power curve to forecast wind power.

Some studies classified forecasting models into two categories namely, point forecasting methods and probabilistic forecasting methods. Unlike the first methods which predict the amount of wind generation at a future point and generate only a single value of the future power, probabilistic forecasting methods attempt to estimate the predictive distributions in the forms of intervals. (Sideratos & Hatziargyriou, 2012) stated that traditional point wind power predictions are fairly low in average and probabilistic wind power models can give much information on uncertainties which are highly useful to power system operators. Prediction of the wind power and wind speed has the same principle, therefore the following section discusses aforementioned models applied in both wind power and wind speed forecasting.

2.4.1 Statistical Methods

Statistic methods attempt to establish the relation between input data, i.e., weather predictions, wind speed, and the generated wind power by statistical analysis of the historical time series. In fact, the objective of such methods is to formulate the pattern of the measurements. In other words, such methods use historical data observed at site under investigation to mathematically formulate the problems. Statistical methods have better performance in short-term prediction than in long-term prediction and are immensely popular due to the simplicity in the structure, and moderate cost (J. Chen, Zeng, Zhou, Du, & Lu, 2018; Pearre & Swan, 2018).

Persistence method (Zhang, 2012), known as the naïve predictor, is the simplest model, typically used as a benchmark for other models, which simply assumes the future value of the wind power is equal to the previously measured one.

ARMA models are the most popular in the time-series based methods to forecast the future value of wind power or wind speed. This model was applied by (Torres, Garcia, De Blas, & De Francisco, 2005) to forecast the hourly average wind speed up to 10-hour ahead using nine years of historical data and the comparative results show 20 % improvement compared to persistence model.

(Erdem & Shi, 2011) suggested four methods based on autoregressive moving average (ARMA) model to forecast the wind speed and direction tuple. In wind speed forecasting the model based on the traditional-linked ARMA outperformed than others, while, in wind direction forecasting model based on the decomposition of wind speed into the lateral and longitudinal components illustrated better accuracy.

(Liu, Shi, & Erdem, 2010) introduced a novel wind power prediction method, namely modified Taylor Kriging (MTK). In this research, the distance formula between two points in covariance function was modified and results indicate that the proposed model outperformed ARIMA.

Times series-based methods include such as Kalman filter (Louka et al., 2008), Autoregressive (Huang & Chalabi, 1995), autoregressive integrated moving average (ARIMA) (Sfetsos, 2002), fractional-ARIMA (Kavasseri & Seetharaman, 2009), seasonal-ARIMA (Meng), and limited-ARIMA (LARIMA) (P. Chen, Pedersen, Bak-Jensen, & Chen, 2010). Classical times series analysis (traditional statistical technique), however, cannot always establish the nonlinear relationship between input and output data because they are based on a linear regression model.

The advent of the matching learning algorithms such as artificial neural networks (Flores, Tapia, & Tapia, 2005; Shuhui Li, Wunsch, O'Hair, & Giesselmann, 2001), Support Vector Machines (SVMs) (Niu, Wang, & Wu, 2010), fuzzy logic (FL) (G. Sideratos & N. Hatziargyriou, 2007), and genetic programming (GP) (Lee & Tong, 2011) caught researcher's attraction in recent years. Since these methods learn the wind power behavior from the previous data, they do not need any previous modeling of the wind power. Indeed, unlike conventional statistical methods, the learning approaches are capable of representing the nonlinear characteristics of the inputs. ANN, for example, a model inspired by the structure of the neural processing in the brain, known as 'gray box' was used in a large number of recent publications due to the self-learning, easy implementation, and efficiency in the modeling the nonlinear relationships (Ak, Li, Vitelli, & Zio, 2018).

Short-horizon prediction of wind speed and wind power using data-driven approaches was presented by (Kusiak & Zhang, 2010). Exponential smoothing, NN, boosting tree, random forest, support vector machine, and the k-nearest neighbor were applied to forecast wind speed based on 10-s data collected and the most accurate model is selected for up to 60s prediction. Then, three models were investigated to forecast wind power as a function of wind speed.

(Kusiak, Zheng, & Song, 2009c) employed several data mining algorithms to build time series models for forecasting the power of the wind farm at different time horizon. According to comparative analysis, SVM outperformed other data mining approaches in wind power and wind speed prediction over 10-min to 60-min ahead, while multilayer perceptron provided most accurate result in prediction of wind power at 1-hour to 4-hour ahead. In this study, boosting tree algorithm and the wrapper approach using genetic search algorithm were used in order to obtain the best predictors.

Long-term wind power and wind speed forecasting (hourly prediction up to 72-h) for a wind farm on the Greek island of Crete was produced by (Barbounis, Theocharis, Alexiadis, & Dokopoulos, 2006). Wind speed and wind direction were used as inputs of three types of local recurrent neural network namely, the infinite impulse response multilayer perceptron (IIR-MLP), the local activation feedback multilayer network (LAF-MLN), and the diagonal recurrent neural network (RNN).

A detailed comparison between ARMA and NNs was provided by (De Giorgi, Ficarella, & Tarantino, 2011) to establish a power forecasting model for a wind farm in Southern Italy, a country with unstable meteorological conditions. Prediction models were made over 1, 3, 6, 12-hour into the future. Wind speed and wind power over five years were employed to define a prediction model, however, wind direction and temperate were not taken into consideration. Five types of NNs were considered in this research: three networks, MLFF, MLP, ElMAN based on only one input (the hourly average wind power) and other two networks, MLP, and ElMAN based on two inputs (the hourly average wind power and wind speed data).

(Palomares-Salas, De la Rosa, Ramiro, Melgar, Agüera, et al., 2009) Compared ARIMA model and back-propagation neural network (NNT) in terms of the Pearson's correlation coefficient, root mean square error (RMSE), and Index Of Agreement

(IOA). Based on the obtained results, ARIMA yielded better accuracy than the NNT in short-time horizon forecasting.

An algorithm based on fuzzy logic was presented by (Damousis, Alexiadis, Theocharis, & Dokopoulos, 2004) to forecast wind speed and produced power at a wind park. Wind speed and wind direction collected from neighboring meteorological stations were averaged within a time range of 15 to 30 minutes. The effectiveness of the model was compared to the persistence model.

A new model based on Resource Allocating Network (RAN) was suggested by (Han, Romero, & Yao, 2015) to forecast wind power within prediction length of up to one day for different types of wind turbines with different capacities. In the data processing stage, a phase space reconstruction (PSR) was utilized to convert the observation into space vector, thereby studying and detecting the dynamic structure of the wind. Then, principal component analysis (PCA) was applied to eliminate noise and redundancies through mapping the original signal to a new higher dimension space. The performance of the model was compared to persistence and new reference (NR) models.

An application of SVM for wind speed prediction based on mean daily wind speed data over a 12-year period was presented by (M. A. Mohandes, Halawani, Rehman, & Hussain, 2004). The obtained results were better than the one obtained by MLP in terms of RMSE and MAE.

2.4.2 Physical Methods

Physical models use numerical weather prediction (NWP) and take into consideration some factors including the local surface roughness, the effects of obstacles and terrain, temperature, and pressure. NWP usually uses complicated equations to present the behavior of the atmosphere. Indeed, NWP employs governing equations such as conservation of mass, conservation of energy, conservation of momentum, and thermodynamics laws. The forecast accuracy of the NWP method depends heavily on initial data. Therefore, the gap in the initial data due to scarcity of observations in the remote areas, such as a mountain, reduces the accuracy of the NWP models (Hoolohan, Tomlin, & Cockerill, 2018). In addition, the manipulating immense amount of data sets and solving complicated equations necessitates utilizing supercomputers in these models. Despite these difficulties, the NWP model is considered one the best prediction methods in long-term forecasting. Although (N. Chen, Qian, Nabney, & Meng, 2014) applied NWP for short-term prediction, mostly this model is used in long-term horizon because it generates more accurate prediction in long-term than does in short-term.

(Landberg, 1999) suggested a model to predict the power generated by a wind farm connected to the grid. The model was based on wind speed prediction from the high-resolution limited area model (HIRLAM) of the Danish Meteorological Institute. The verification of the model was performed using 10-min average data over one-year period collected from 17 wind farms with total capacity of 35.7 MW. The obtained results were compared to those of the persistence model.

(Lazić, Pejanović, & Živković, 2010) examined the application of a regional numerical weather prediction Eta model in wind forecasting for the wind power plants. The Eta model was evaluated using two different resolutions; one was larger area and coarse resolution (22*22 km) and the other was smaller area and finer resolution (3.5*3.5). The obtained wind from Eta model at 10 m level was compared to the observed wind from the surface station and the wind turbine at 10 m. Further investigation was conducted between the obtained wind from Eta and the one from a wind turbine at 38, 54, 75, and 96 m.

(Jimenez, Durante, Lange, Kreutzer, & Tambke, 2007) drew a comparison between two WNPs approaches, mesoscale meteorological model MM5 and the wind resource assessment program WAsP (Wind Atlas Analysis and Application Program), to estimate the wind resource over the German Bight in the North Sea.

A short-term wind power forecasting on a real case study in southern Italy was investigated using two different ensemble prediction models, the Ensemble Prediction System (EPS) in use at the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Limited-area Ensemble Prediction System (LEPS) developed within Consortium for Small-scale Modelling (COSMO) (Alessandrini, Sperati, & Pinson, 2013). The results revealed that the higher resolution model (COSMO-LEPS) generated slightly better performance notably from 27 to 48-hour ahead.

A new combination of particle swarm optimization (PSO) algorithm and Type-2 fuzzy neural network (T2FNN) known as T2FNN-PSO was conducted by (Sharifian, Ghadi, Ghavidel, Li, & Zhang, 2018). The performance of the method was verified using data from an online supervisory control and data acquisition (SCADA) system and the numerical weather prediction (NWP) for a medium-term wind power prediction.

An application of Kalman filtering in numerical prediction of wind speed was presented by (Louka et al., 2008). To forecast wind speed, two limited-area atmospheric models with different horizontal resolution were applied. The results demonstrated that Kalman filter not only can eliminate the systematic forecast errors in NWP but also reduced the CPU time. (Nielsen, Nielsen, Madsen, Pindado, & Marti, 2007) also attempted to reduce the error of the forecast model by using several NWP.

A comparison study was carried out by (Carvalho, Rocha, Gómez-Gesteira, & Santos, 2014) to evaluate the performance of the weather research forecast (WRF) mesoscale model in the wind simulation and wind energy estimation by different initial and boundary forcing conditions. Six different WRF simulations were applied and their obtained results were compared to the observed wind data from thirteen wind measuring stations located in Portugal.

2.4.3 Hybrid Methods

Combined prediction approaches can improve the accuracy of the forecasting taking advantage of individual prediction methods which have, of course, different performance depending on the forecast horizon, the data sets, and their capability to map nonlinear relationships. Combined techniques have great importance in a wide range of application area due to the fact that individual method has a good performance only in a certain situation and thus different models have to be evaluated and tested to obtain the best performance. Combined techniques not only settle the time-consuming drawback of individual approaches but also provide great superiorities over the individual models. In general, hybrid methods are a combination of different models such as a combination of statistical and physical models (Katinas, Gecevicius, & Marciukaitis, 2018), a combination of any evolutionary algorithms with statistical (Mahmoud, Dong, & Ma, 2018) or physical approaches or a combination of different models with different time horizon either (Leng et al., 2018; Nourani Esfetang & Kazemzadeh, 2018).

A novel statistical method using the artificial neural network and fuzzy logic for wind power prediction was developed by (G. Sideratos & N. D. Hatziargyriou, 2007). To make the best use of NWP and improve its inaccuracy, firstly a self-organized map was applied to divide input data according to the magnitude of the wind speed, into three classifications: small, medium, and large. Secondly, a separate radial basis function (RBF) network was used to provide an initial prediction for each class.

Subsequently, a fuzzy logic model was employed to indicate the quality of the predicted wind speed by NWPs. In summary, the main contribution of the proposed model was the optimum use of NWPs available based on fuzzy logic rules.

An integration of kernel principal component analysis (KPCA) and evolutionary optimized local general regression neural network (EOLGRNN) was derived by (Elattar, 2014). The function of KCPA in the proposed model was to construct the phase space and to overcome the drawback of the conventional time series reconstruction technique. Since the smoothing parameter had a great impact on the performance of the GRNN, an evolutionary algorithm (EA) was applied to obtain the optimum value of the smoothing parameter. In terms of accuracy, the proposed model outperformed other models, but it sacrificed operation cost and speed nevertheless. EA, indeed, slowed the training procedure down.

Two different models were investigated by (Peng, Liu, & Yang, 2013) for short-term wind power forecasting; one was a hybrid strategy based on ANN (a statistical model) and the physical models and the other one was individual ANN with nine neurons in the hidden layer. The performance of two models was evaluated using real data (wind speed, wind direction, and temperature) collected from a wind farm with the total rated power of 50 MW. The numerical results revealed that combined model was considerably more accurate than individual ANN in terms of MAE and NRMSE, however, individual ANN was able to make a faster prediction.

An effective wind power forecasting model based on Bayesian clustering by dynamics (BCD) and support vector regression (VSR) was developed by (Fan et al., 2009). To reduce the sensitivity of the model to the input variables, all the inputs data were divided into two groups: wind speed and wind generation data were used as inputs of SVRs, while wind direction and humidity in addition to wind speed and wind generation data were considered inputs of BCD. Temperature and pressure were not included in data sets because the author found these inputs do not produce any improvements in the results but, conversely, slowing down the learning process. In order to select only input variables with significant influence on the output, the autocorrelation and cross-correlation between different attributes (inputs) were analyzed. However, they are linear and might not catch the nonlinear dependencies between input variables.

A hybrid method, combining ANN and wavelet transform (WT) was introduced by (Catalão, Pousinho, & Mendes, 2011) for short-term wind power forecasting in Portugal. The WT was applied to decompose the original wind power series into different scale components. Due to the filtering effect of WV, each scale component illustrated better behaviors than the original series. Then, NN attempted to forecast the future value of each scale component. At the end, an inverse WT reconstructed the future behavior of the wind power series. In this research, NNWT was compared to NN, ARIMA and persistence model.

(Pousinho, Mendes, & Catalão, 2011) developed a model for short-term wind power prediction based on 15 min ahead data. The proposed model was based on ANFIS and particle swarm optimization (PSO). The PSO played an essential role in tuning the parameters of the membership function resulting in great improvement in the performance of the ANFIS. Historical wind power data are used as inputs of ANFIS, however, other exogenous variables such as temperature and wind direction were not taken into account. The comparative analysis verified the superiority of the proposed model over ARIMA, NN, NNWT, wavelet-neuro-fuzzy (WNF) and persistence models.

A comparative study based on nearest neighbor search (NNS) and ANN was conducted by (Jursa & Rohrig, 2008) to predict the generated power in 10 wind farms in

Germany. Two population-based optimization algorithms namely, PSO and differential evolution (DE) were used to select the most relevant input variables. Input variables consisted of numerical weather predicted data (NWP) and the observed power of the wind farms over 3.5 years. The NN and NNS using evolutionary algorithms demonstrated better forecasting accuracy as compared to when their inputs were selected manually. For further evaluation, both models were compared to persistence model.

A hybrid model comprising of WT, RBF neural network, imperialist search algorithm (ICA), and MLP was presented by (Aghajani, Kazemzadeh, & Ebrahimi, 2016) to predict the output power of the wind farm located in the southern part of Alberta, Canada. As a primary predictor, RBF neural network forecasts the wind power. Then, continuous WT (CWT) was applied to filter the input data consisting of the output of RBF neural network and other exogenous variables, i.e. wind direction, wind speed, humidity, temperature, and wind direction. In the next step, MLP attempted to predict the future value of each decomposed signal by WT. MLP, defined as the main predictor in this study, use three different learning algorithms: Levenberg–Marquardt (LM), Broyden–Fletcher–Goldfarb–Shanno (BFGS), and Bayesian regularization (BR). Subsequently, ICA aimed to optimize the weights of neural networks. Finally, predicted signal of wind power was reconstructed through an inverse WT. The proposed model (RBF+ HNN + WT + ICA) outperforms RBF + HNN + WT + PSO, RBF + HNN + WT, RBF + HNN, and HNN + PSO + WT.

A short term wind speed and wind power prediction model using a multi-layer feed-forward neural network (MFNN) and simultaneous perturbation stochastic approximation (SPSA) algorithm were developed by (Hong, Chang, & Chiu, 2010). Three different structures of MFNN were considered in this research. A cascaded structure where forecasted wind speed served as an input for power prediction; a parallel structure where all inputs (wind speed and power) were applied simultaneously and a separated MFNNs in which forecast engines were separately operated. The inputs of all three models were modeled by fuzzy numbers. SPSA attempted to train the MFNN and improved its convergence. In order to illustrate the efficiency of the developed model, it was compared to other models such as MFNNs with the back-propagation algorithm, ARMA, time-interval averaging approach (TIAA), and persistence method.

(Amjady, Keynia, & Zareipour, 2011a) suggested a new strategy for short-term prediction of wind power based on ridgelet neural network (RNN). Moreover, a new differential evolution algorithm (NDE) with an efficient crossover operator and selection mechanism was introduced to train the developed forecast engine. In order to remove the unimportant inputs and obtain the candidates with a high share of information, a feature selection technique called mutual information (MI) was used. Analysis results showed that the proposed model yields better accuracy than the persistence model, ARIMA, radial basis function (RBF) neural network and multi-layer perceptron (MLP) neural network trained by Levenberg–Marquardt (LM) learning algorithm. Further evaluations revealed the efficiency of NDE in comparison with the simulated annealing (SA), genetic algorithm (GA), particle swarm optimization (PSO) and classical DE.

2.4.4 Indirect Wind Power Prediction Methods

This approach comprises two processes: first, the wind speed as most effective parameter among exogenous variables is needed. In the real world application, these data are typically purchased from some meteorological institutes such as Deutscher Wetterdiens or can be provided by some services, for instance, WAsP (Wind Atlas

Analysis and Application Program), HIRLAM (high-resolution limited area model) (Croonenbroeck & Dahl, 2014). However, these data are quite expensive notably when they are required at high resolution. Thus, using the stochastic or/and hybrid technique to predict the wind speed at wind farms are cost effective. Subsequently, the forecasted wind speed, in particular time horizon, is utilized to obtain wind power prediction through the power curve of the operational wind turbine machine.

Fractional-ARIMA or f-ARIMA as a special case of ARIMA model was proposed to forecast wind speed over 24- and 48-hour future time intervals (Kavasseri & Seetharaman, 2009). Unlike traditional ARMA and ARIMA, f-ARIMA was well suited to capture long range correlation existing in the wind speed. The developed model was validated using real data collected from four wind parks in North Dakota. The f-ARIMA model provided better forecast accuracy as compared to the persistence model, ARIMA and neural network in terms of the daily mean error (DME), the variance, and the square root of the forecast mean square error. Lastly, wind power was estimated through forecasted wind speed and the power curve of a NEG Micon 750 kW wind turbine operating in the site under investigation.

An application of Grey rolling model (GM) to forecast hourly wind speed and wind power was suggested by (El-Fouly et al., 2007). Three versions of GM model were applied so as to improve the drawback of the traditional GM. The adaptive GM produced better results as compared to the persistence and traditional GM model but only for some intervals. The modified GM model managed to reduce the overshoot, however, a good agreement was not reached between the observed and actual values. The averaged GM model illustrated most satisfactory than others. Lastly, V66-1.65 MW wind turbine power curve manufactured by VESTAS Company was employed to predict the hourly wind power.

A statistical wind power forecasting model based on WT and SVM was applied by (Zeng & Qiao, 2012). In order to improve the ability of original SVM, a new kernel function was proposed which is able to switch between radial basis function (RBF) kernel and a Mexhat kernel. The developed WPP model included three steps. Data normalization was performed in the first step. Then, the original wind speed series was decomposed using wavelets and subsequently, SVM was used to forecast the future behavior of the wind speed. Wind power was obtained from the wind turbine's wind-power curve (Vestas V-90 3-MW) according to the forecasted value of the wind speed. The WSVM model was compared to the persistence and RBF-SVM.

(Senjyu, Yona, Urasaki, & Funabashi, 2006) presented the output power prediction of wind turbines based on wind speed forecasting using the recurrent neural network (RNN). In RNN, unlike FNN, there was a feedback structure transmitting information from hidden layer to input layer. The optimum number of hidden neurons was determined by trial and error. The efficacy of the proposed model over the FNN was verified in terms of MAPE.

A hybrid of neural network and genetic algorithm were applied for short term wind speed forecasting by (Senjyu et al., 2006). Originally, the weights and bias of NN were optimized by BP in conjunction with gradient descent search, however, there is no guarantee the global minimum can be reached. In the developed model, GA was applied to overcome this drawback and to improve the forecast accuracy of NN. In fact, GA attempted to obtain the best weight vector of BPNN. The performance of the model was compared to BP and momentum BP. In the last step, predicted wind speed was converted to wind power through manufacturer power curves.

(Focken et al., 2002) compared the statistical analysis of the power forecasting error of an ensemble wind park at a single site. The accuracy of the power prediction was evaluated within 6-, 12-, 18-, 24-, 36-, and 48-hour ahead. Numerical weather prediction model provided by German weather service was used for wind speed prediction. Lastly, the wind speed prediction results were translated into wind power forecasts through the power curve of the wind turbine.

2.4.5 Direct Wind Power Prediction Methods

Unlike indirect prediction techniques, these models use meteorological data such as temperature, pressure, wind direction, humidity, wind speed and historical information of wind energy output as inputs to predict the future value of the wind power generation. In these models, wind speed might be forecasted. Some of the proposed forecast engines require the wind speed at the hour of prediction as the most important input, thus in this case, similar to indirect prediction models, firstly wind speed should be forecasted, however wind power curve is not required. In other words, input variables directly predict the wind power generation.

(Mabel & Fernandez, 2008) applied NN to forecast wind power output of wind parks in Muppandal, India with an installed capacity approximately 1000 MW. Three variables wind speed, humidity, and recorded wind energy output over a period of three years were used in this study.

A comprehensive review on wind power forecasting using data mining approaches was provided by (Colak, Sagiroglu, & Yesilbudak, 2012). This research specified which forecasting model produces better accuracy for very short term, short term, medium term, and long-term forecasting of wind power. As a result of this survey, ANFIS demonstrated better performance in very short, ANN in short-term, and MLP in medium and long-term wind energy predictions. Moreover, it was found that in direct prediction models wind power and wind speed are mainly used as inputs, however, pressure and temperature have a profound impact on wind energy.

NWP data including wind speed, wind direction and temperature over more than a year was applied by (Xu et al., 2015) for short-term wind forecasting. Since NWP data often contains several inaccuracies, a new outlier detection mechanism based on data-mining technique was introduced to identify the improper NWP data. Indeed, the k-means algorithm was applied to partition observations into several clusters and then detecting the abnormal NWP data. The number of clusters was determined based on Bayes information criterion (BIC). Finally, an NN was used to predict wind power based on adjusted NWP data.

To predict the generation of a wind farm an ensemble model including WT, NN, feature selection technique, and partial least square regression (PLSR) was developed by (Song Li, Wang, & Goel, 2015). Only wind power and wind speed were used as input variables to the forecast engine while temperature and humidity were not considered. The most informative input variables were selected by a conditional mutual information technique. In order to capture more precise information from the wind power data, a WT was applied to break up given data into several components. To select the best setting of the wavelet, features of 12 different mother wavelets were considered in wavelet architecture. Each individual then was connected to a feedforward NNs with Levenberg–Marquardt learning algorithm to predict the future value. Each individual output was assigned to a different weight and combined using PLSR.

A hybrid neuro-fuzzy wind power forecasting system for wind power generation was presented by (Saleh, Moustafa, Abo-Al-Ez, & Abdullah, 2016). The prediction system used historical data as well as measured data by a wireless sensor network (WSN) including wind speed, wind power, and temperature. To obtain the optimum number of fuzzy rules, the clustering of the dataset was conducted by using fuzzy c-means (FCM). The efficacy of the model was evaluated by RMSE and relative error.

(Amjady, Keynia, & Zareipour, 2011b) composed an efficient feature selection technique and forecast engine based on NWP data including humidity, temperature, wind direction, and wind speed provided by Canadian Meteorological Centre and U.S. National Centers for Environmental Prediction. To obtain the most relevant input variables with minimum redundancy, two-stage feature selection technique based on mutual information was employed. Based on the selected inputs, NN attempted to map nonlinear relation between inputs and output power. The proposed enhanced PSO (EPSO) was added to NN for assisting in escaping from a local minimum. The numerical experiments evaluated the effectiveness of the proposed FS technique in a comparative manner. Furthermore, the superiority of the developed forecast engine was compared to hybrid NN (HNN), modified HNN (MHNN) using GA, MHNN using DE, and MHNN using PSO.

2.5 Prediction Time Horizon

In addition to methodology, wind power forecasting techniques can also be categorised in terms of time horizon. Overall wind power prediction approaches can be divided into four groups based on time scales: very short-term, short-term, medium-term, and long-term wind power forecasting. The first time scale provides forecast from few seconds to 30 minutes ahead and is used for load tracking and wind turbine control as well as electricity market and real-time grid operations. Short term wind forecasting techniques are developed for pre-load sharing and economic load dispatch planning and include the forecasting from 30-min to 6-hour. Mainly statistical techniques for example ANNs, are employed for short-term wind power prediction due to the fact that models based on NWP deal with long-running calculations. Similarly, combined prediction techniques are inefficient in short-term prediction due to longer operation time as compared to the individual approaches. Medium-term predictions are utilized for energy trading and power system management and ranging from 6-hour to one-day ahead. The last time scale refers to one-day to one-week ahead forecasting and used for the maintenance planning and repair of the wind turbine machines. It is noted that medium and long-term predictions are typically based on NWPs.

2.6 Wind Speed Distribution

A comprehensive assessment of wind energy regime is of fundamental importance that should be carried out prior to implementing any wind energy project. It is due to the fact that the cost of producing energy is heavily dependent on the wind energy at site.

The wind energy in any site is strongly influenced by the wind speed, which has a cubic relationship to the generated wind power. Knowledge of wind characteristic not only assists the estimation of the future revenue and income but also facilitates a due turbine design selection for the chosen location. The probability distribution of wind speed is the important information required for the assessment of wind energy potential. For this reason, in order to accurately analyze the characteristics of wind speed frequency distribution, a variety of PDFs was proposed.

(Brano, Orioli, Ciulla, & Culotta, 2011) applied Weibull, Rayleigh, Lognormal, Gamma, and inverse Gaussian to estimate the wind speed frequency distributions. (Zhou, Erdem, Li, & Shi, 2010) not only compared Weibull and Rayleigh distributions, but also Gamma, Lognormal, Inverse Gaussian, and MEP-PDF with six goodness-of-fit statistics. (Jamil, Parsa, & Majidi, 1995) used the two-parameter Weibull probability distribution function to find out the wind energy density from the statistical data of wind speed measurements. The Weibull and lognormal models were proposed for fitting wind speed distribution by (Garcia, Torres, Prieto, & De Francisco, 1998). (Lun & Lam, 2000) employed a two-parameter Weibull function to describe the wind speed frequency distribution for a given set of wind data for three different locations. (Celik, 2004) presented a statistical analysis of wind power density based on the Weibull and

Rayleigh models. Monthly wind energy production was calculated using the Weibull-representative wind data and it was shown that the Weibull function estimates accurately the wind energy output. In order to develop a more accurate method for estimation of wind speed characteristics, (S. Akdağ, Bagiorgas, & Mihalakakou, 2010) analyzed the characteristics of wind speed data using typical two-parameter Weibull wind speed distribution and the two-component mixture Weibull distribution, involving five parameters. (Fan et al., 2009; Jiang, Wang, Wu, & Geng, 2016) drew a comparison between the gamma, Weibull, and Rayleigh probability density functions and it was found that all three wind speed distributions give a good agreement to the recorded wind speed data, though Weibull distribution was recommended due to its simplicity. Based on the literature, Weibull and the Rayleigh functions are the most two common distributions in which the Rayleigh distribution is a subset of the Weibull distribution and Weibull distribution is a special case of two-parameter Gamma function. The reason for their popularity is that the estimation of their parameters is not difficult. The Weibull is a two-parameter distribution while the Rayleigh has only one parameter. This makes the Weibull more versatile and the Rayleigh simpler to use. In detail, the reasons why two-parameter Weibull is widely applied and accepted among other distribution functions is that it fits the wind distribution very well; it has a flexible structure, varying according to the shape parameter of the distribution; it provides easy determination of parameters; the number of parameters is few; and once the parameters for a certain height are determined, the wind data for various heights can be calculated using the already determined parameters (S. A. Akdağ & Dinler, 2009; Chang, 2011; Dursun & Alboyaci, 2011).

2.7 Wind Turbine Power Curve

The wind turbine power curve (WTPC) represents the relationship between the electrical output power of the wind turbine and hub height wind speed. Wind energy assessment is one of the prime objectives for modeling of the WTPC. Wind energy assessment is an important procedure usually performed by wind farm developers to estimate the future energy generation of a wind farm. A comprehensive and accurate assessment is a requisite for the successful development of the wind farm. A WTPC can greatly facilitate estimation of wind power that can be generated over a period of time if the wind speed data in site is available. An accurate wind turbine power curve model, as a prediction tool plays a key role in electricity market wherein wind power participants must pay penalty if they underestimate the future wind energy produced.. WTPC models, furthermore, aid wind farm developers to choose the most appropriate wind turbines which would deliver optimum efficiency and improved performance. It is of great important that a WTPC model can serve as a very effective performance monitoring tool (Kusiak et al., 2009b).

As illustrated in Figure 2.2, past studies on wind turbine power curve modeling are broadly divided into two categories: models based on fundamental equations of power available in wind and models based on the concept of the power curve of the wind turbine. The former depends on various parameters, i.e. wind speed, the rotational speed of the turbine, turbine angle of attack and pitch angle, mechanical transmission efficiency, generator efficiency and so on. Due to the interdependence of these parameters and their variation with change in climatic conditions, and type of components used as well, using these models is not only cumbersome but also do not give accurate results.

To approximate wind turbine power curve an expression was developed by (Ashok, 2007) which is highly dependent on air density, power coefficient of wind generator, wind turbine rotor swept area, as well as the wind turbine and generator efficiency. In the study, however, some aspects were ignored such as variation in the value of wind turbine and generator efficiency with speed, variation in the value of air density with changing weather condition, and variation in the value of electrical output power of the turbine for various wind speed ranges. To improve the performance of the previous study, another expression with an additional term, the efficiency of AC/DC converter, was employed by (Nelson et al., 2006). Although the variation of the turbine output power with wind speed was considered, neglition on air value variation with weather condition remained. Moreover, the efficiency of the generator and mechanical transmissions was not taken into account. (Kolhe, Agbossou, Hamelin, & Bose, 2003) provided a formula to analyze the performance of the wind turbine generator combined with the photovoltaic array in a stand-alone renewable energy system. In that research, neither the efficiency of the generators was considered nor the variation in the value of air density and output power with wind speed was discussed. Another expression to calculate the power captured by the wind turbine was developed by (Thanaa, Eskander, & El-Hagry, 2006). In this model, however, variation in the value of air density with changing weather condition was considered, but yet the efficiency of turbine and generator was disregarded.

Several wind power prediction models are critically analyzed by (Thapar, Agnihotri, & Sethi, 2011) and it was concluded that the behavior of wind turbines might not correctly replicated by the models based on the fundamental equation of wind power, whereas the performance of a wind turbine can be conveniently modeled by the concept of the power curve. In fact, the power curve represents the amount of generated electrical power by a wind turbine at a specific wind speed, without the technical details of the components of the wind turbine. In other words, unlike models based on fundamental equations of power available in wind, models using the concept of power curve do not require details of various parameters of the wind generator. These models can be broadly divided into three categories: models based on a presumed shape of the power curve, models based on actual power curve supplied by the manufacturer, and models based on the empirical power curve.

There is a general presumption in models based on a presumed shape of the power curve that the power curve of the wind turbine follows a typical shape. Hence, a set of expressions and equations were developed to forecast wind turbine output for a various range of wind speed. It is well noted that a particular set of characteristic equation may not guarantee to replicate the behavior of all types of turbines. This is due to the fact that different types of wind turbines are different in shape of power curve depending on the control method and strategy, design, and power capacity. Accordingly, models based on presumed shape of the power curve, though are simple to use, they do not yield promising results.

(HX Yang, Lu, & Burnett, 2003; Hongxing Yang et al., 2009) presented a simple model to predict the performance of the wind turbine in which it was assumed that the relationship between wind speed and output power of the wind turbine in the medium wind speed is linear. Though the proposed model was simple, it does not yield accurate result because considering the power curve, the generated power seldom linearly increases with wind speed. The model based on cubic law also showed an inaccurate result. This is often rooted in dependency of the term η, turbine efficiency, to weather condition and turbine blade parameters (Chedid et al., 1998). The model using Weibull's parameters, too, could not perfectly map the non-linear relationship between the wind speed and the wind turbine's output power (Karaki, Chedid, & Ramadan, 1999; Lu, Yang, & Burnett, 2002; Powell, 1981).

Models based on actual power curve supplied by manufacturers attempt to accurately predict the output electrical power of the wind turbine through mathematical expressions. To do so researchers employed various curve fitting techniques to find the best fit to the actual power curve of the individual wind turbine. The characteristic equation of wind turbine was obtained by fitting its actual power curve based on the least square method (Ai, Yang, Shen, & Liao, 2003). A model based on three binomial expressions was presented and the accuracy of the model on different wind turbines with different shapes of power curve was investigated. All turbines considered, however, were only medium-sized wind turbines, ranging from 260 to 335 kW. (Diaf, Diaf, Belhamel, Haddadi, & Louche, 2007; Hocaoğlu, Gerek, & Kurban, 2009) produced a model to optimize the size of the hybrid wind/PV system. In this context, the output power of the wind turbine was estimated by cubic spline interpolation based on the power curve data provided by manufacturers (theoretical wind power curve). In these researches, the accuracy of the developed model was only evaluated on a particular type of wind turbine and its performance on different type of wind turbines was not disclosed. In general, the main drawback of the models based on the actual power curve is that they focus on the power curve supplied by the manufacturers which usually consider the wind turbines operate under ideal condition. It is well noted that in practice the empirical power curve does not perfectly match theoretical power curve which resulting in inaccuracy of the models based on the actual power curve. In fact, power curves supplied by manufacturers are tested at specific air density which might be different with the air density of the site locations. In addition, disregard of the local turbulence as well as the wear and tear of wind generator components result in indisputable discrepancy between the theoretical and the empirical power curves.

The last group, the models based on the empirical power curve, includes parametric and nonparametric techniques. The former which is developed in this research, emplyes mathematical equations to approximate the wind turbine power curve model. Logistic four-parameter (4-PL) (Kusiak et al., 2009b), Polynomial expression (Giorsetto & Utsurogi, 1983), and logistic five-parameter (5-PL) (Lydia et al., 2013) are examples of this approach. While the latter attempts to find a relationship between the wind power and the wind speed. Neural network (NN) models were employed by (Marvuglia & Messineo, 2012) to monitor a wind farm's power curve with assumption of normal operating conditions ,however it is not easy to backtrack the output of NN due to its black-box nature.

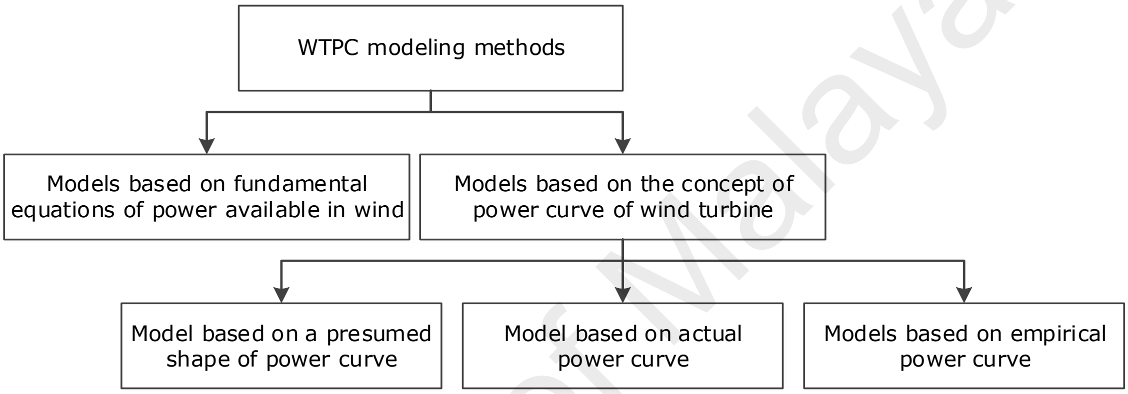


Figure 2.2: Classification of WTPC modeling

CHAPTER 3: WIND TURBINE POWER CURVE MODELING

3.1 Introduction

In this Chapter, a new wind turbine power curve (WTPC) modeling is introduced. The proposed WTPC is applied later in the indirect wind power prediction in Chapter 4. Furthermore, in Chapter 4, the detailed performance evaluation of direct power predictions are also investigated and then compared with the indirect power prediction. Figure 3.1 illustrates the main research works in this study representing WTPC and Feature Selection Technique (FS) covered in Chapter 3 and Chapter 4 respectively.

In the first part of this chapter a new parametric model, called modified hyperbolic tangent (MHTan), is introduced to approximate the empirical wind turbine power curve. To obtain the coefficients of the proposed wind-power curve modeling two methods with different procedures are employed. One attempts to find the best coefficients of the developed wind-power modeling by minimizing the squared residuals. To do so, three heuristic optimization algorithms are employed. In the other one, wind speed distribution and maximum likelihood are used to estimate the coefficients. Therefore, firstly several methods are developed to obtain the parameters of the wind speed distribution. Then, power density function of the wind turbine based on the power-curve model and wind speed distribution is derived. The coefficients of the derived formula are indeed the unknown parameters of the proposed power-curve modeling. The performance of the methods applied for the estimation of wind speed distribution is analyzed based on both actual and simulated data. Data from a real wind farm are used to validate the performance of the MHTan in approximating the wind power curve model. The result of the MHTan is compared with other methods from the literature. To pose a challenge to the MHTan, three different sets of data based on three wind turbines with different shape of power curve and size of power are used. In addition, the applicability of MHTan in online monitoring is presented in the last part of this chapter.

The programming code was written in Matlab and executed on a personal computer with Intel Pentium 2.66 GHz processor and 4 GB RAM.

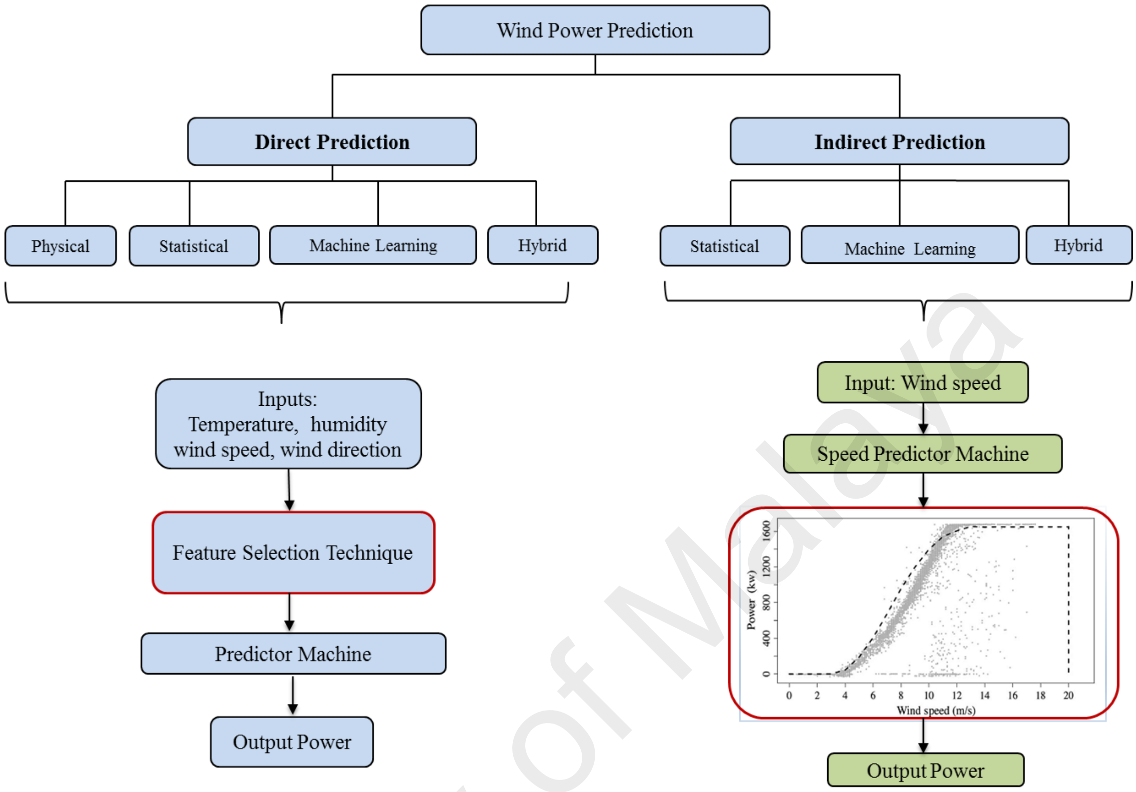


Figure 3.1: The Proposed WTPC and Feature Selection Technique (FS)

3.2 Development of Wind Turbine Power Curve Model

The strong relationship between the wind speed and wind power can be depicted by wind turbine power curve. The Turbine performance as a function of wind speed is illustrated in Figure 3.2 for three different types of wind turbines. Cut-in speed, V_c, is the minimum wind speed needed for a turbine to generate the power. The nonlinear relationship between the wind speed and power continues until nominal wind speed, V_r . The third region on the wind power curve depends on the control strategy of the wind turbine. Above the nominal wind speed, in pitch-based turbines, the output power is maintained, while in yaw-based and stall-based wind turbines the output power declines. Cut-out (V_f) speed is the maximum speed in which the power extracted from the turbine remains constant. To avoid any damage to the turbine, , it will be shut down above the cut-out speed (Jafarian & Ranjbar, 2010).

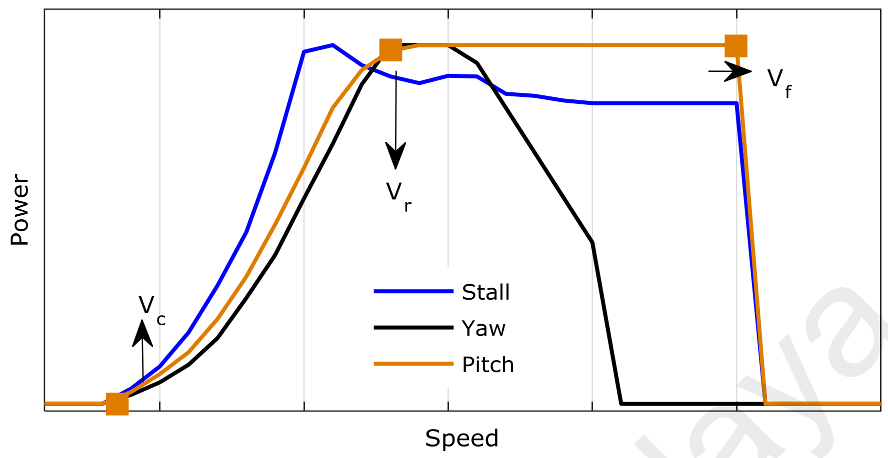


Figure 3.2: Output power curves based on of three different turbine control strategy

The power curve shape can be modeled by modified hyperbolic tangent (MHTan). (Tuev VI, 2009) employed MHTan to estimate the current–voltage characteristics of field-effect transistors. The MHTan as a special S-shaped function based on the hyperbolic tangent, is proposed in this study to build an accurate model for wind turbine power curve and its accuracy is compared to the prior parametric models. The MHTan function is given by:

$$y_e = \frac{a_1 \exp(a_2 x) - a_3 \exp(-a_4 x)}{a_5 \exp(a_6 x) + a_7 \exp(a_8 x)} + a_9 \qquad (3.1)$$

where $\theta = (a_1, a_2, \dots, a_9)$ is a vector of unknown parameter of MHTan which determines its shape. When parameters $a_1, \dots, a_8$ are set to 1 and a_9 to 0 the MHTan reduces to hyperbolic tangent, whereas for $a_1 = a_2 = a_3 = a_4 = a_5 = a_7$ and $a_6 = a_8 = a_9 = 0$, Eq. (3.1) becomes equal to hyperbolic sine. Unlike the parameters of the 5- PL (a_3 and a_5), there is no constraint on the range of the parameters in MHTan. Two different methods are applied in this study to obtain the parameters of MHTan as shown in Figure 3.3; one based on least square error (LSE) and the other based on maximum likelihood estimation. Both methods are explained in detail in the following sections. The developed model can be used as a reference for monitoring the performance of wind turbines which is especially useful in offshore wind farms because of accessibility and oversight issues. Wind turbine performance monitoring substantially reduces the operation and maintenance cost (Leung & Yang, 2012).

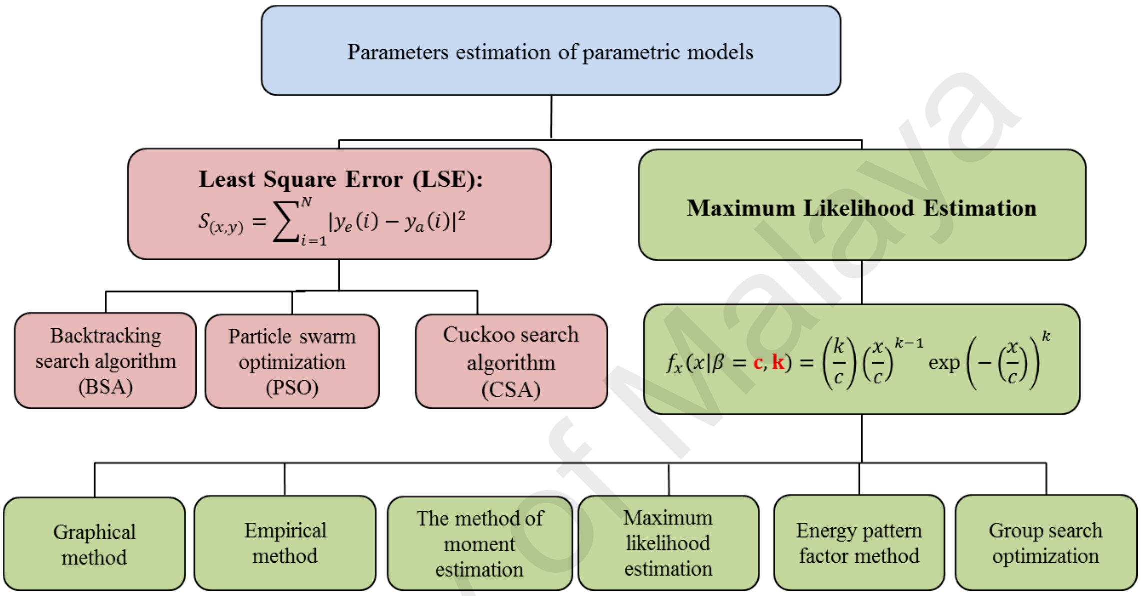


Figure 3.3: Applied methods to obtain the coefficient of the parametric WTPC models

3.3 Parameter Estimation of the WTPC Model Using Least Square Error (LSE)

One of the statistical methods that can be used to determine the coefficients of a wind turbine power curve model is the LSE method. LSE technique can be applied to compute the unknown parameters of wind turbine power curve modeling method. The optimum fitting curve is presumed and obtainable by minimizing the deviations square summation from a data set. Given that N pairs of observations defined as $[x_1\ y_1], \ldots, [x_N\ y_N]$, x and y represent the independent and dependent variables respectively. The coefficients of Eq. (3.1) can be estimated by minimizing the cost function Eq. (3.2).

$$S_{(x,y)} = \sum_{i=1}^{N} {r_i}^2 = \sum_{i=1}^{N} [y_e(i) - y_a(i)]^2 \qquad (3.2)$$

where y_a and y_e present the actual power and the power estimation from the parametric models respectively. The $\hat{\theta}$ with L vector coefficient membership number can be calculated as the following (for each individual parametric model):

$$\hat{\theta} = \underset{a_1, \dots, a_L}{\text{argmin}} \quad S_{(x,y)} \left((x_1, y_1), \dots, (x_N, y_N) | a_1, \dots, a_L\right) \qquad (3.3)$$

By utilizing metaheuristic optimization techniques, the parametric model vector coefficient is obtainable. These techniques are widely implemented to achieve optimum solutions. In this research, cuckoo search algorithm (CSA) (Ikeda & Ooka, 2015; Nguyen, Vo, & Truong, 2014; Piechocki, Ambroziak, Palkowski, & Redlarski, 2014)**,** backtracking search algorithm (BSA) (Civicioglu, 2013), and particle swarm optimization (PSO) (AlRashidi & El-Naggar, 2010; Fong, Yuen, Chow, & Leung, 2010) are applied for minimization of Eq. (3.2) which are explained in details in the section below.

3.3.1 Backtracking Search Algorithm (BSA)

One of the finest methods to overcome optimization problems is developed by (Civicioglu, 2013) and technically known as BSA. It has a simple structure with only one parameter control, which is suitable to overcome even multimodal optimization based problems. Unlike other methods or techniques, the BSA's performance is not too sensitive to its parameter control and moreover it is free from premature convergence and huge computational cost. To efficiently discover research domain, BSA uses mutation and crossover operators in the algorithm. It is distinguishable from other techniques such as genetic algorithm, evolutionary programming, and etc. In addition, BSA has an outstanding memory that drives the search direction, depending on the previous or earlier generations. The illustration chart of BSA is shown in Figure 3.4. It has five main steps; namely initialization, selection-I, mutation, crossover, and selection-II.

Step 1 (initialization): the parameters, $\mathbf{Pop} = [\mathbf{X_1}\ \mathbf{X_2}\ \mathbf{X_3} \cdots \mathbf{X_{nPop}}]'$ and $\mathbf{X_i} = [\mathrm{x_{i1}} \cdots \mathrm{x_{ij}} \cdots \mathrm{x_{i\,D}}]'$, which represent population and individual of each group are initialized. The parameter i and j are defined as individual and element numbers. Eq. (3.4) generates the initial population, which includes nPop individuals. The D optimization variables, is included in each individual.

$$Pop_{ij} = x_{ij} \sim U(low_j, up_j) \tag{3.4}$$

where:

i: defined as individual, *i* = (1,2,…, nPop)

j: defined as optimization variable, *j* = (1,2,3,…, D)

low_j and up_j: lower and upper limits of variable j

U: uniform distribution function

x_{ij}: is the *j*th element of the *i*th individual as the member of the population

Step 2 (selection-I): this step generates historical population or **histPop. Pop** and **histPop** are similar in the size. Furthermore, x_{ij} element in **histPop** is actually a x_{ij} counterpart in **Pop**. Eq. (3.5) is initialized at the beginning to produce **histPop**. After that, reclassification of historical population is done via the "if-then" order by matching up two numbers (randomly) which is defined as *a* and *b*, based on the Eq. (3.6). Lastly, for completion, individual numbers of **histPop** is randomly modified via Eq. (3.7). A random shuffling function, acting as a permuting function is applied in the mentioned equation. The search direction for each iteration is determined using the **histPop**.

$$histPop_{i,j} \sim U(low_{i,j}, up_{i,j}) \tag{3.5}$$

$$if\ a < b \left| \begin{array}{l} .g \quad hh \quad \rightarrow \mathbf{histPop} = \mathbf{Pop} \\ a, b \sim U(0,1) \end{array} \right. \tag{3.6}$$

$$\mathbf{histPop} = premuting(\mathbf{histPop}) \tag{3.7}$$

Step 3 (mutation): mutant, which technically defined as trial population initial form, is produced from the mutation process via Eq. (3.8). **Pop** is subtracted from **histPop** to determine the direction of the search, while the search direction amplitude is controlled by the function F. For standard normal distribution, the F value is set to 3.rand (rand stands for random).

$$Mutant = \mathbf{Pop} + F.(\mathbf{histPop} - \mathbf{Pop}) \tag{3.8}$$

Step 4 (crossover): this step known as crossover process is used to finalize the Mutant (explained in step 3). The process involves changing Mutant to the final population trial **TR** via crossover handler. Firstly, TR value is set according to the Mutant. A nPop rows and D columns forming a binary matrix (map) are randomly generated. The matrix 'map' each row is individually relevant. A BSA parameter called 'mixrate' is a single control parameter that manages the string of elements of any individual within the crossover process. This control parameter (ranges from 0-100% of D elements) determines the maximum number of elements in each row of the binary matrix "map" to be equal to 1. Two techniques can be employed in this crossover process; the first one is to engage random individual elements and the second one is to choose the maximum mixrate individual elements for manipulation in crossover procedure. According to the strategy, firstly the binary matrix (map) is produced and then TR elements with the corresponding value of 1 in the matrix (map) will be manipulated. For this case, TR elements are placed so that it will be equal to the **Pop** relevant elements, or can be simplified as map_{ij}=1 then TR_{ij}=Pop_{ij}.

Step 5 (boundary control): it may be seen that certain boundary limits have been violated by some individual elements. In this case, Eq. (3.4) can be optionally set to its upper or lower limits. It may be seen that certain boundary limits have been violated by some individual elements. In this case, Eq. (3.4) can be optionally set to its upper or lower limits Eq. (3.4).

Step 6 (selection-II): in this step, to get the optimum fitness value, a comparison is drawn between each individual of **TR** and **Pop**. Upon comparison, individuals of **Pop** will be updated.

Step 7 (BSA's control parameter and stopping condition): in the crossover procedure, 'mixrate' is the only BSA's control parameter throughout optimization process. The optimization of BSa is not excessively sensitive (range between 0-100%), but fine-tuning is required to obtain the best optimal performance. In addition, to control the optimization procedure, a stopping or halt condition is required. Usually, the process will choose the maximum number of iterations to define the stopping/halting condition.

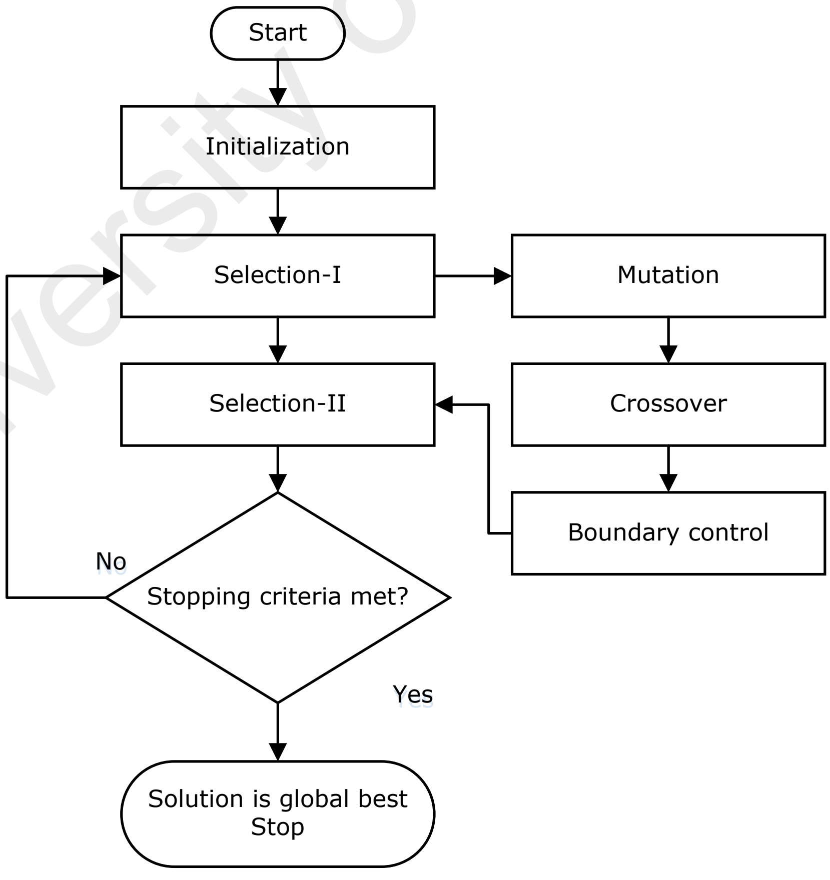


Figure 3.4: Flowchart of backtracking search algorithm

3.3.2 Particle Swarm Optimization (PSO)

The particle swarm optimization algorithm, a stochastic population-based metaheuristic, is first introduced by (Kennedy & Eberhart, 1995; Shi & Eberhart, 1998). The idea of this evolutionary algorithm came from the observation of the behavior of natural organisms, e.g. birds and fish, to find food. PSO algorithm operates with a swarm of the particle, in which, each particle changes its direction due to either leader's command or environmental impacts. In the PSO, population (swarm) consists of a lot of particles and each particle as a solution to a particular problem in the search space that contains two characteristics: its own position and velocity. The particles would move from the current position based on the speed and random direction.

The position, indeed, represents the current values in the solution. Particles accelerate towards the particle having the best fitness value. In other words, the leader has a great effect on the speed and direction of each particle. Thus, each particle shifts its current position to its new position. In each iteration, particles approach a better position and would stop moving as the swarm reaches to the best position. Due to a simple principle, high efficiency, and high search speed PSO has been widely used in multi-objective optimization, mode identification, signal processing, damage identification, etc.

As presented in Figure 3.5, basically PSO algorithm comprises three main steps: generating positions and velocities of particles (solutions), velocity update, and position update which are described as follows:

Step 1 (initialization): the population of particles is randomly generated within the minimum and maximum limits of the parameters of the particular function, for example, modified hyperbolic tangent. In a D-dimensional search space, the complete swarm consists of nPop particles which are represented as a matrix as follows:

$$\mathbf{Pop} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_j, \dots \boldsymbol{X}_{nPop}] \tag{3.9}$$

where $\boldsymbol{X}_j = [x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{Dj}]$ is the position vector of the *j*th particle and it is one of the solutions to parameter estimation. In parallel, the velocity vector is randomly initialized and then distributed uniformly in the search space within appropriate limits. The speed vector of *j*th particle is defines as:

$$\boldsymbol{V}_j = [\boldsymbol{v}_{1j}, \boldsymbol{v}_{2j}, \dots, v_{ij} \dots, \boldsymbol{v}_{Dj}] \tag{3.10}$$

Step 2 (evaluating the particles): the fitness of each particle $\boldsymbol{X}_j$ is evaluated based on the assigned objective function Eq. (3.2).

Step 3(initialization of the best position): there are two key factors in PSO strategy called, Pbest and Gbest. The former is the most optimist position of each particle which is defined as $\boldsymbol{Pbest}_j^t = [pbest_{1j}^t, pbest_{2j}^t, \dots, pbest_{ij}^t \dots, pbest_{Dj}^t]$, while the Gbest is the best position out of all Pbest and defined as $\boldsymbol{Gbest}^t = [gbest_1^t, gbest_2^t, \dots, gbest_i^t,$ $\dots, gbest_D^t]$. In the first iteration, the initial starting point of each individual is taken as its Pbest.

Step 4 (movement of particles): in the iterative search process, position and speed of each particle are updated based on global and individual best positions. The change of the speed of each individual can be expressed as follows:

$$v_j^{t+1} = w.v_j^t + \rho_1.c_1\left(pbest_{ij}^t - x_{ij}^t\right) + \rho_2.c_2(gbest_i^t - x_{ij}^t) \tag{3.11}$$

where, $i \in \{1,2,\dots,D\}$, $j \in \{1,2,..,P\}$, c_1 and c_2 are the accelerated factors which control the speed of individuals, ρ_1 and ρ_2 are the random numbers uniformly distributed $U(0,1)$ in each iteration t, and w is the inertia weight. After updating the speed, the particle moves from current position to the new position which is formulated in Eq. (3.12).

$$x_{ij}^{t+1} = x_{ij}^{t} + v_{j}^{t+1} \tag{3.12}$$

Step 5 (updating the best positions): every particle is able to remember information to determine its best position in the search space according to the fitness value and has the ability to share information so as to obtain the best position found by the swarm and save it. The new position of each individual is evaluated through Eq. (3.12) and then Pbest is updated as follows:

$$\begin{cases} \boldsymbol{Pbest}_j = \boldsymbol{X}_j + \boldsymbol{V}_j & if \;\; F'_j < F_j \\ \boldsymbol{Pbest}_j = \boldsymbol{Pbest}_j & if \;\; F'_j \geq F_j \end{cases} \tag{3.13}$$

where F' is the evaluated objective value of the jth particle at the new position.

Step 6 (stopping condition): steps 2-5 will be repeated until the number of iterations exceeds the maximum or the desired value of fitness function is reached.

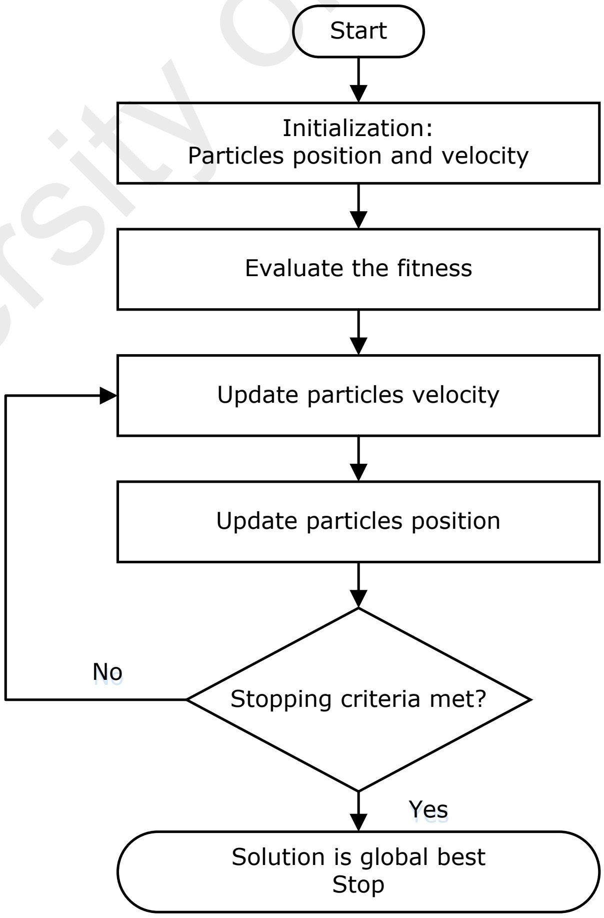


Figure 3.5: Flowchart of particle swarm optimization

3.3.3 Cuckoo Search Algorithm (CSA)

Cuckoo search is a metaheuristic algorithm, inspired by nature, and developed by (X.-S. Yang & Deb, 2009) for solving optimization problems. In general, it is based on cuckoo breeding parasitic behavior and Lévy flights. The former is characterized by a process in which cuckoos lay their eggs in the nests of other host birds where the host bird just laid its own eggs, so as to increase their egg hatching rate. The host takes care of the eggs presuming that the eggs are its own, however, if the host bird identifies the alien egg, it certainly either throws them away or abandon the nest and builds a new nest at a different location. The more foreign eggs similar to the host bird's egg, the higher chance they have to develop (next generation) and become a mature cuckoo. These eggs are usually high-quality eggs that are close to optimum value. The cuckoo search algorithm uses this behavior as a model, traversing the search space to find optimal solutions.

The use of Lévy flights is of vital importance to the cuckoo search. Lévy flights are the forward steps taken by birds in search of food. These steps are randomly taken and depend on the current location and the transition probability to the next location. The direction of random forward steps follows a probability density function which is modeled mathematically. This random walk is derived from Lévy distribution with an infinite variance and mean.

There are three principal rules for the CSA described as follows:

1. Each cuckoo among the fixed number of available host nests randomly chooses a nest in which it lays its egg.
2. The most favorable nest with higher quality of egg, representing better solutions preserved for the next generation.

3. The number of available host nests is fixed, and a host bird can reveal the alien eggs with a probability $P_a \in [0,1]$.

Based on these three basic rules, CSA is applied to estimate the unknown parameters of parametric models. As shown in Figure 3.6, the steps involved in CSA are as follow:

Step 1 (initialization of host nests): the algorithm starts with an initial population of nPop hosts which is represented by $\mathbf{Pop} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{nPop}].$, $i = 1,2, \ldots nPop$. The population uses D-dimension vector restricted within minimum and maximum limits as given in Eqs. (3.14) and (3.15).

$$\boldsymbol{X}_{imin} = [x_{1min}, x_{2min}, \ldots, x_{Dmin}] \quad (3.14)$$

$$\boldsymbol{X}_{imax} = [x_{1max}, x_{2max}, \ldots, x_{Dmax}] \quad (3.15)$$

Therefore, $\boldsymbol{j}$th component of the ith population is initialized as follows:

$$x_{ij} = x_j^{min} + rand_1. (x_j^{max} - x_j^{min}) \quad (3.16)$$

where $j=1,2,\ldots,D$ and $rand_1$ is a uniformly distributed random number between 0 and 1.

Step 2 (generation of new solution via Lévy flights): generating the new nests includes two stages, generating new nests including Lévy flights and replacement of a fraction of eggs. With the exception of the best nest, all other nests are replaced based on the quality of new cuckoo eggs which are generated by Lévy flights from their position. In this stage, Mantegna's algorithm is used to calculate the optimal path for the Lévy flights. The new solution by each nest is calculated as follows:

$$X_i^{new} = Xbest_i + \vartheta * rand_2 * \Delta X_i^{new} \quad (3.17)$$

where $\vartheta > 0$ is the step size parameter which are used to adjust the convergence rate of the algorithm; *rand₂* is a normally distributed stochastic number; and the increased value ΔX_i^{new} is determined by:

$$\Delta X_i^{new} = \frac{rand_x}{\left|rand_y\right|^{\frac{1}{\kappa}}} * \frac{\sigma_x(\kappa)}{\sigma_y(\kappa)} * (Xbest_i - Gbest_i) \tag{3.18}$$

where $rand_x$ and $rand_y$ are two normally distributed stochastic variables with the standard deviation $\sigma_x(\kappa)$ and $\sigma_y(\kappa)$ given by:

$$\sigma_x(\kappa) = \left[\frac{\Gamma(1+\kappa) * \sin(\frac{\kappa\pi}{2})}{\Gamma\left(\frac{1+\kappa}{2}\right) * \kappa * 2^{(\kappa-1)/2}}\right]^{1/\kappa} \tag{3.19}$$

$$\sigma_y(\kappa) = 1 \tag{3.20}$$

where κ is the distribution factor $0.3 \leq \kappa \ll 1.99$ and Γ is the gamma distribution function.

Step 3 (alien egg discovery and randomization): as mentioned earlier, there is the possibility that the foreign egg detected by the host bird and it then may be thrown away out of the nest. Similar to Lévy flight, the discovery action of alien egg in the nest of host bird with the probability of P_a also creates a new solution for the problem.

$$\boldsymbol{X_i^{Discovery}} = \boldsymbol{Xbest_i} + \boldsymbol{K} * \Delta\boldsymbol{X_i^{Discovery}} \tag{3.21}$$

where K is the updated coefficient determined based on the probability of a host bird to discover a foreign egg in its nest:

$$K = \begin{cases} 1 & if\ rand_3 < P_a \\ 0 & otherwise \end{cases} \tag{3.22}$$

the increased value, $\Delta\boldsymbol{X_i^{Discovery}}$ is obtained by:

$$\Delta \boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{Discovery}} = rand_4 * \left[rand_{p1}(Xbest_i) - rand_{p2}\left(Xbest_{p2}\right)\right] \tag{3.23}$$

where $rand_3$ and $rand_4$ are the distributed random numbers in [0,1] and $rand_{p1}$ $(Xbest_i)$, $randp_2(Xbest_i)$ are the random perturbation for positions of the nests in $Xbest_i$.

Step 4 (termination criteria): the algorithm is terminated as the number of iterations reaches the predefined value or a desired value of fitness function is reached.

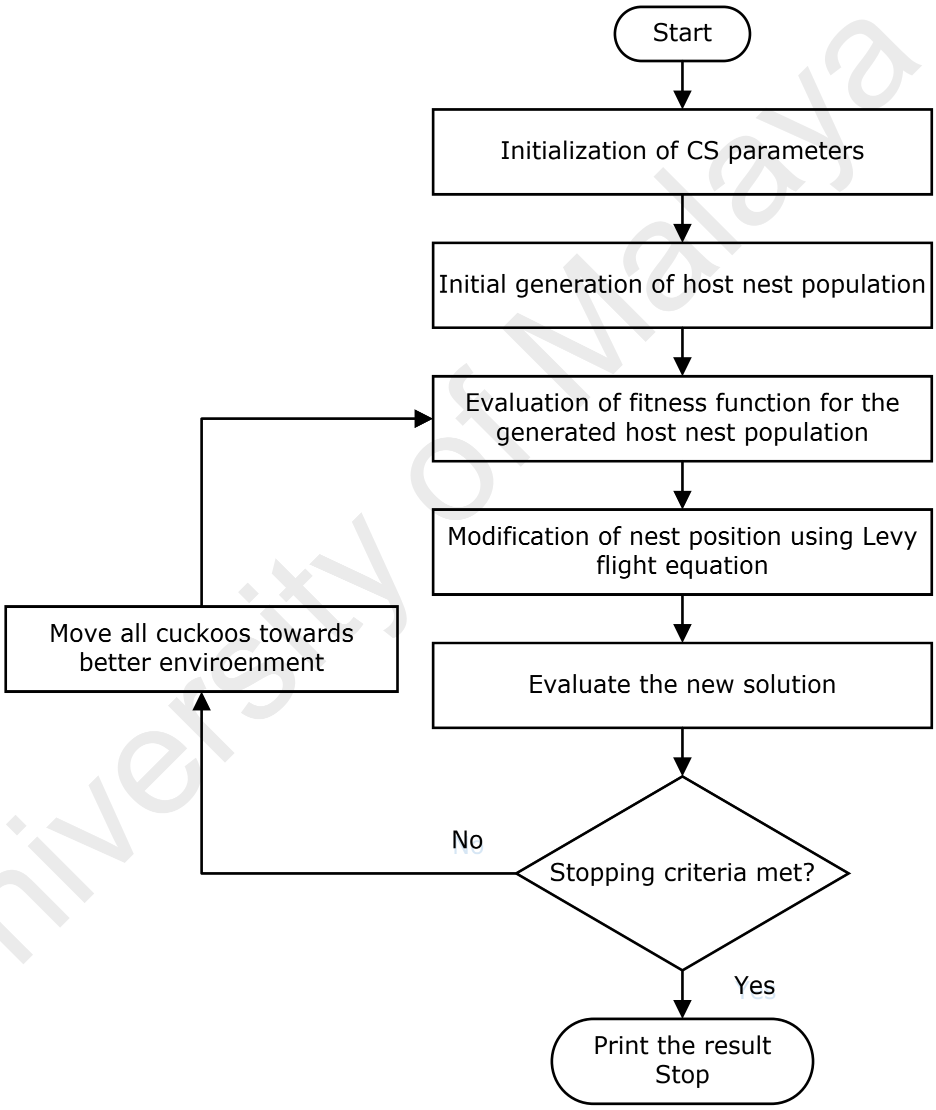


Figure 3.6: Flowchart of cuckoo search algorithm

3.4 Parameter Estimation of the WTPC Model through Maximum Likelihood Estimation

An alternative approach to estimate the coefficients of the MHTan is by means of maximum likelihood estimation. To do so, firstly a proper distribution function representing wind speed behavior must be selected and it's parameters then must be obtained.. Here, Weibull distribution density function is chosen as per literature. Then several techniques are applied to estimate the parameters of the selected wind speed distribution. In the next step, a new formula representing probability density function (PDF) of WTPC based on Weibull and MHTan is derived. The unknown parameter of the derived PDF is indeed the unknown parameter of MHTan.

The probability density function of the Weibull distribution is given by (Ucar & Balo, 2009):

$$f_x(x|\beta = c, k) = \left(\frac{k}{c}\right)\left(\frac{x}{c}\right)^{k-1} \exp\left(-\left(\frac{x}{c}\right)\right)^k \tag{3.24}$$

where *x* is the wind speed, *c* is the scale factor, with units equal to the wind speed units, and *k* is the dimensionless shape factor. The higher value of the scale parameter represents that wind speed is higher, while the value of shape parameter remarkably influences the shape of the distribution curve and indicates the wind stability. In other words, parameter *c* varies according to the average wind speed and parameter *k* indicating the wind frequency. If the shape parameter is exactly 1, then the distribution is called Exponential distribution. If *k* =2, the probability distribution follows Rayleigh PDF and would be similar to normal distribution if *k*=3.5. Generally, manufacturers of wind turbines provide the standard performance of their machines using the Rayleigh distribution.

The Weibull cumulative distribution function (CDF) can be obtained by calculating the integral of the PDF presented by Eq. (3.24). The cumulative function relevant to a two-parameter Weibull distribution is expressed as follows (Ohunakin, Adaramola, & Oyewola, 2011):

$$F(x) = 1 - \exp\left[-\left(\frac{x}{c}\right)^{k}\right] \quad (3.25)$$

3.4.1 Estimation of Weibull Parameters

It is well-known that parameter estimation has significant effects on the success of Weibull distribution for wind energy applications. Hence, in this section, in order to accurately estimate the parameters of Weibull PDF, six numerical methods, namely, graphical method, empirical method, moment method, maximum likelihood estimation, energy pattern factor method and a metaheuristic algorithm are employed.

3.4.1.1 Graphical method (GM)

GM based on the concept of least squares is implemented to fit a straight line to wind data, where time-series data must be ranked in ascending order. Taking the double natural logarithmic transformation of the Weibull CDF, given in Eq. (3.25), the following equation can be developed (Justus, Hargraves, Mikhail, & Graber, 1978):

$$ln\left[-ln\left(1 - F(x)\right)\right] = k\, ln(x) - k\, ln(c) \quad (3.26)$$

This equation is considered as the simple linear model: $y = \beta_0 + \beta_1 t$ where $y = ln\left[-ln\left(1 - F(x)\right)\right]$, $t = ln(x)$, $\beta_0 = -k\, ln(c)$ and $\beta_1 = k$. Estimation of the slope $\beta_1$and intercept β_0 can be obtained through the ordinary least square estimation or by manual drawing. Finally, the parameters of the Weibull distribution are estimated as follows:

$$c = Exp\left(-\frac{\beta_0}{\beta_1}\right) \;\; \& \;\; k = \beta_1 \quad (3.27)$$

3.4.1.2 Empirical method (EM)

The empirical method could be considered as a special case of the moment method, where the Weibull shape parameter is estimated by (Rocha, de Sousa, de Andrade, & da Silva, 2012):

$$k = \left(\frac{\sigma}{\bar{x}}\right)^{-1.086} \tag{3.28}$$

and then the scale parameter is obtained as:

$$c = \left(\frac{\bar{x}}{\Gamma\left(1 + \frac{1}{k}\right)}\right) \tag{3.29}$$

3.4.1.3 The method of moment estimation (MM)

One of the techniques to estimate population parameters is MM. It works by substituting the theoretical moments (distribution of population) from the mean and variance sample (by using sample moments) in place of wind speed theoretical moments. Supposing that random variable x is the wind speed observations and unknown parameters of the PDF is defined as β, the wind speed PDF will be $f(x, \beta)$. When utilizing MM, the existence of the origin moment (mean sample of wind speed), which is used to estimate the whole distribution expectation, μ is applied. On the other hand, the central moment or the second moment (variance sample of wind speed) is utilized in estimation of the variance population, σ^2. The formula of the above discussion can be express as follow s (Usta, 2016):

$$\bar{x} = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.30}$$

$$\sigma = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2} \tag{3.31}$$

where n is the number of the wind speed observations. The shape and scale parameters can be obtained when μ and σ satisfy the following formula:

$$\left(\frac{\sigma}{\mu}\right)^2 = \frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1 \tag{3.32}$$

The scale parameter c can be calculated by the following expression:

$$c = \frac{\bar{x}}{\Gamma\left(1+\frac{1}{k}\right)} \tag{3.33}$$

where $\Gamma(.)$ is the Gamma function expressed by:

$$\Gamma(\mathrm{a}) = \int_0^\infty t^{a-1} e^{-t} dt \tag{3.34}$$

3.4.1.4 Maximum likelihood Method (MLM)

To statistically estimate the parameters of wind speed distribution model, MLM technique can be employed. Its basic principle lies on the maximizing the sample wind speed probability, based on the model distribution interpretation. In MLM, indeed, by maximizing the likelihood function, unknown parameters estimation can be achieved. This estimation is depending on the underlying population distribution. Note that, before MLM is used, advanced population distribution must be known.

Assume n size random sample $x_1, x_2, \dots, x_n$ are drawn from a set M with PDF $f_x(x,\beta)$, where the unknown parameter is defined as β. In this case, $x_1, x_2, \dots x_n$ probability function is given by:

$$L = \prod_{i=1}^{n} f_{x_i}(x_i, \beta) \tag{3.35}$$

and the MLM of vector β is the value of β that maximizes L or similarly the L algorithm can be written as follows:

$$\frac{d\log L}{d\beta} = 0 \tag{3.36}$$

Based on the Eq. (3.24)**,** the function of likelihood can be written as (Arslan, Bulut, & Yavuz, 2014)**:**

$$L(x_1, x_2, \dots, x_n, k, c) = \prod_{i=1}^{n} \left(\frac{k}{c}\right)\left(\frac{x}{c}\right)^{k-1} \exp\left(-\left(\frac{x}{c}\right)\right)^{k} \tag{3.37}$$

By substituting Eq. (3.37) into Eq. (3.36) and differentiating with respect to k and c, the following expression can be obtained.

$$\frac{\partial \ln L}{\partial k} = \frac{n}{k} + \sum_{i=1}^{n} \ln x_i - \frac{1}{c}\sum_{i=1}^{n} x_i^k \ln x_i = 0 \tag{3.38}$$

$$\frac{\partial \ln L}{\partial c} = \frac{-n}{c} + \frac{1}{c^2}\sum_{i=1}^{n} x_i^k = 0 \tag{3.39}$$

The elimination of variable c between Eqs. (3.38) and (3.39) and a number of simplifications resulting in the following expression:

$$\frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{k} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0 \tag{3.40}$$

This can be accomplished by Newton-Raphson method which can be written in the following form:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{3.41}$$

where,

$$f(k) = \frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k} - \frac{1}{k} - \frac{1}{n}\sum_{i=1}^{n} ln\, x_i \tag{3.42}$$

and,

$$f'(k) = \sum_{i=1}^{n} x_i^k (\ln x_i)^2 - \frac{1}{k^2} \sum_{i=1}^{n} x_i^k (k \ln x_i - 1) - (\frac{1}{n} \sum_{i=1}^{n} \ln x_i)(\sum_{i=1}^{n} x_i^k \ln x_i) \tag{3.43}$$

The iterative equation of k is gained by the following equation:

$$k = \left[\frac{\sum_{i=1}^{n} [x_i^k \, ln(x_i)]}{\sum_{i=1}^{n} x_i^k} - \frac{1}{n} \sum_{i=1}^{n} ln \, x_i \right]^{-1} \tag{3.44}$$

once k is determined, c can be estimated as

$$c = \left[\frac{\sum_i^n x_i^k}{n} \right]^{1/k} \tag{3.45}$$

3.4.1.5 Energy pattern factor method (EPFM)

In order to estimate the shape k and scale c parameters with this PDM, first, the energy pattern factor E_{pf} is calculated as (Raichle & Carson, 2009):

$$E_{pf} = \frac{\overline{x^3}}{(\bar{x})^3} \tag{3.46}$$

where $\bar{x}$ is the sample mean of wind speed, and $\overline{x^3}$ is the sample mean of wind speed cubes. Then the two parameters are estimated using the following equations:

$$k = 1 + \frac{3.69}{\left(E_{pf}\right)^2} \tag{3.47}$$

Scale parameter is estimated using Eq. (3.33).

3.4.1.6 Group search optimization (GSO)

Having been inspired by animal behavior, notably, animal searching (foraging) behavior, (He, Wu, & Saunders, 2009) proposed GSO primarily for continuous optimization problems. The population of the GSO algorithm is called a *group* and each individual in the population is called a *member*. In general, there are two foraging strategies within groups: one is to search for food, and the other is to join resources (food) discovered by others. The former is called producing and the latter is called joining or scrounging. In GSO, however, a group comprises three kinds of members: producers, scroungers, and dispersed members. The latter carries out random walk to avoid any entrapments in local minima. In GSO, unlike other algorithms, producers and scroungers do not perform based on information sharing (IS) model at which it is generally assumed that foragers search for their own resources while searching for other opportunities to join. Indeed, in GSO forages are assumed to use either producing or joining strategies exclusively. The flowchart of GSO is illustrated in Figure 3.7, while the detailed procedures are as follows:

Step1: all of the M group of members are randomly generated in a D dimensional search space so that the ith $(i = 1,2, \dots, M)$ member at the tth iteration has a current position $\boldsymbol{X_i^t} \in \mathbb{R}^D$, head angle $\boldsymbol{\varphi}_i^t = (\boldsymbol{\varphi}_{i_1}^t, \dots, \boldsymbol{\varphi}_{i_{(D-1)}}^t) \in \mathbb{R}^{D-1}$with search direction of $\mathbf{D}_i^t(\boldsymbol{\varphi}_i^t) = (d_{i_1}^t, \dots, d_{i_D}^t) \in \mathbb{R}^n$ which can be calculated from $\boldsymbol{\varphi}_i^t$ through polar to Cartesian coordinate transformation as follows:

$$
\begin{gathered}
d_{i_1}^t = \prod_{q=1}^{D-1} \cos(\boldsymbol{\varphi}_{i_q}^t) \\
d_{i_j}^t = \sin\left(\boldsymbol{\varphi}_{i_{(j-1)}}^t\right) . \prod_{q=j}^{D-1} \cos(\boldsymbol{\varphi}_{i_q}^t) \qquad (j = 2, \dots, D-1) \\
d_{i_D}^t = \sin(\boldsymbol{\varphi}_{i_{(D-1)}}^t)
\end{gathered}
\tag{3.48}
$$

Step 2: at each iteration, a group member which is located in the better position is selected as the producer. Scanning is of vital importance to the search orientation wherein GSO, it is accomplished through the visual mechanism and employed by the producer. The scanning field of vision is characterized by the maximum pursuit angle

$\beta_{max} \in \mathbb{R}^1$ and maximum pursuit distance $l_{max} \in \mathbb{R}^1$ as illustrated in a 3-D space in Figure 3.8.

The producer scans searching space such that covers zero point and its lateral area including one point in the left-hand side hypercube and the other in the right-hand side hypercube, which expressed as follows respectively:

$$\begin{aligned} \mathbf{X}_z &= \mathbf{X}_p^t + r_1 l_{max} \mathbf{D}_p^t(\boldsymbol{\varphi}^t) \\ \mathbf{X}_l &= \mathbf{X}_p^t + r_1 l_{max} \mathbf{D}_p^t(\boldsymbol{\varphi}^t - \mathbf{r}_2 \beta_{max}/2) \\ \mathbf{X}_r &= \mathbf{X}_p^t + r_1 l_{max} \mathbf{D}_p^t(\boldsymbol{\varphi}^t + \mathbf{r}_2 \beta_{max}/2) \end{aligned} \tag{3.49}$$

where $r_1 \in \mathbb{R}^1$and $\mathbf{r}_2 \in \mathbb{R}^{D-1}$ are randomly generated from normal distribution with mean 0 and standard deviation 1 and from uniform distribution in the interval [0,1].

After that, the producer will find the best point by calculating their fitness values. If the best point has a better location as compared to the current position, then the producer will fly to that point and turn its head to a randomly generated angle:

$$\boldsymbol{\varphi}^{t+1} = \boldsymbol{\varphi}^t + \mathbf{r}_2 \gamma_{max} \tag{3.50}$$

where $\gamma_{max} \in \mathbb{R}^1$ is the maximum turning angle.

If the producer cannot find a point which is located in a better area it will return its head angle to zero degree after h interactions,

$$\boldsymbol{\varphi}^{t+h} = \boldsymbol{\varphi}^k \tag{3.51}$$

where $h \in \mathbb{R}^1$ is a constant.

Step 3: a number of group members in each iteration are selected as scroungers that search for any opportunity to join the resources uncovered by the producers. Indeed, they perform based on area copy strategy at which scroungers keep moving across to search immediately around the producer. This behavior can be modeled as a random walk towards the producer as follows:

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + \mathbf{r}_3 \circ (\mathbf{X}_p^t - \mathbf{X}_i^t) \tag{3.52}$$

where $\mathbf{r}_3 \in \mathbb{R}^n$is a random sequence in the range between 0 and 1 generated from uniform distribution. Operator "$\circ$" known as Hadmard product, is an element-by-element product of the two matrices with the same size.

Step 4: the rest of group members are dispersed from their position to assist the group to escape from local minima. In this context, at *t*th iteration, they generate a random head angle through formula a the choose a random distance

$$l_i = h.r_1 l_{max} \tag{3.53}$$

and move to the new position.

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + l_i \mathbf{D}_i^t(\boldsymbol{\varphi}^{t+1}) \tag{3.54}$$

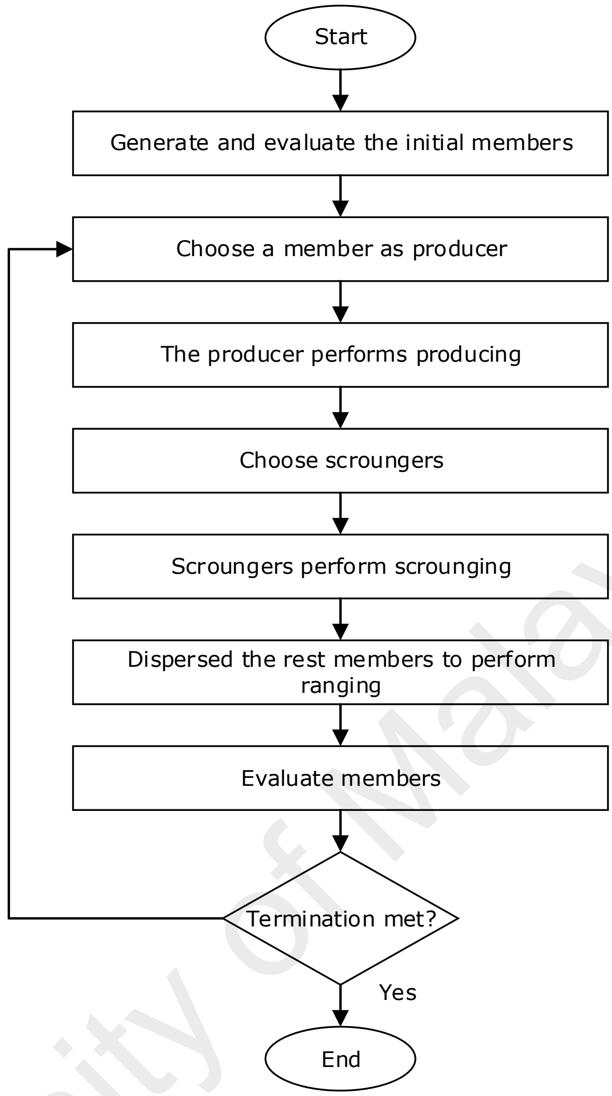


Figure 3.7: Flowchart of group search optimizer algorithm

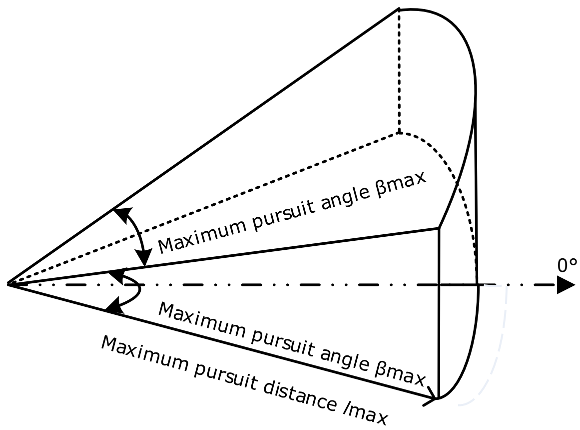


Figure 3.8: Scanning field in 3-D space

3.4.2 The Maximum Likelihood Estimation Based on the Power Distribution

From Weibull distribution of wind speed and turbine power curve model based on MHTan, the pdf of wind turbine output power can be obtained as (Casella & Berger, 2002):

$$f_Y(y) = f_X\big(h^{-1}(y)\big)\left|\frac{dh^{-1}(y)}{dy}\right| \tag{3.55}$$

where x is the wind speed following Weibull distribution denoted as $f_x(.)$, y is the actual power, h represents MHTan model and $f_y(.)$ is an unknown PDF of y. A simplification to Eq. (3.55) has been made by assumption that $a_2 = a_4 = a_6 = a_8 = a$. Thus,

$$h^{-1}(y) = \frac{1}{2a}\log\left(\frac{a_3 + a_7 y - a_7 a_9}{a_1 - a_5 y + a_5 a_9}\right) \tag{3.56}$$

$$\frac{dh^{-1}(y)}{dy} = \frac{a_1 a_7 + a_3 a_5}{2a(a_1 + a_5 a_9 - a_5 y)(a_3 - a_7 a_9 + a_7 y)} \tag{3.57}$$

Based on Eq. (3.24) and differentiating h^{-1} with respect to variable y, Eq. (3.55) leads to:

$$f_Y(y|\theta) = \left|\frac{k(a_1 a_7 + a_3 a_5)}{2ac(a_1 - a_5 y + a_5 a_9)(a_3 + a_7 y - a_7 a_9)}\right|$$
$$\times \exp\left(-\left(\frac{1}{2ac} log\left(\frac{a_3 + a_7 y - a_7 a_9}{a_1 - a_5 y + a_5 a_9}\right)\right)^k\right) \tag{3.58}$$
$$\times \left(\frac{1}{2ac} log\left(\frac{a_3 + a_7 y - a_7 a_9}{a_1 - a_5 y + a_5 a_9}\right)\right)^{k-1}$$

where θ includes the coefficients of MHTan, and k and c are the estimated parameters of the Weibull distribution. With regard to Eq. (3.35), the function of likelihood as an objective function can be written as:

$$L(\theta) = \sum_{i=1}^{N} \log\left(f_y(y_i|\theta)\right) \tag{3.59}$$

The θ , therefore, can be obtained by Eq. (3.60), which can be solved by the three above mentioned optimization techniques in sections 3.3.1 to 3.3.3.

$$\hat{\theta} = \underset{a_1, \dots, a_9}{\operatorname{argmin}} L(\theta) \tag{3.60}$$

3.5 Evaluation Criteria for Weibull Parameters Estimation

To determine the performance of the six aforementioned methods in estimating the parameters of Weibull distribution, the coefficient of determination (R^2), root mean square error (RMSE), Kolmogorov– Smirnov test (K–S), and chi-square (χ^2) are used as the evaluation criteria goodness-of-fit. Goodness-of-fit refers to the fitting degree between the actual and the estimated values, which in this section is the fitting degree between the actual wind speed frequency and the fitted wind speed distribution.

The coefficient of determination (R^2) or squared Pearson correlation coefficient represents the degree of correlation between the estimated wind speed PDF and the experienced probability density. The determination (R^2) is expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_a(i) - y_e(i))^2}{\sum_{i=1}^{N}(y_a(i) - \bar{y}_a(i))^2} \tag{3.61}$$

where N represents the total number of bins which wind speed are divided into, y_a the frequencies of the wind speed received from the measurement, y_e the estimated frequencies calculated from theoretical distribution, $\bar{y}_a$ the average of y_a values.

The root mean squared error (RMSE) is another judgment of accuracy which is calculated as follow:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(y_a(i) - y_e(i))^2\right]^{1/2} \quad (3.62)$$

The chi-square as a non-parametric testis employed to discover the degree of difference between the observed and the expected value. In other words, chi-square goodness-of-fit test determines how well theoretical distribution fits the empirical distribution. In the chi-square goodness-of-fit test, sample data is divided into k bins and the test statistic is defined as:

$$\chi^2 = \sum_{i=1}^{N}\frac{(y_a(i) - y_e(i))^2}{y_a(i)} \quad (3.63)$$

In the statistical test, the critical region for rejection of the hypothesis at the significance level α is,

$$\chi^2 > {\chi^2}_{1-\alpha,df} \quad (3.64)$$

where ${\chi^2}_{1-\alpha,df}$ is the α-quantile of the chi-squared distribution with *df* degrees of freedom.

Kolmogorov–Smirnov test (K–S test) is adopted to decide whether a particular continuous theoretical PDF is suitable to describe the wind speed sample or not. in this context, it compares the cumulative observed frequency distribution with the theoretical distributions then determines that which type of distribution the observations come from. A narrow gap between the empirical CDF and the theoretical distributions affirms that the sample is taken from the theoretical distributions. The K–S test is defined as the max error between two cumulative distribution functions:

$$KS = max|F(x) - O(x)| \quad (3.65)$$

where F(x) and O(x) are the theoretical cumulative distribution function and the empirical cumulative of the actual wind speed. The critical value for the K-S test at 95% confidence level is given by:

$$KS_{0.95} = \frac{1.36}{\sqrt{n}} \tag{3.66}$$

where n is the number of data points in the sampled wind speed data. If KS value exceeds the critical value then one can say that there is a significant difference between the theoretical and the time-series data under the given confidence level.

3.6 Evaluation Criteria for Wind Turbine Power Curve Modeling

To validate the performance of the parametric and nonparametric WTPC modeling methods, two goodness-of-fit indicators, namely mean absolute percentage error (MAPE) and RMSE, are employed. RMSE is defined in Eq. (3.36) while MAPE is defined as follows:

$$\text{MAPE} = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{|y_a(i) - y_e(i)|}{\overline{y(i)}}\right]^{1/2} \tag{3.67}$$

where y_a and y_e represent the observed the estimated power, n is the number of samples, and $\bar{y}$ is the average of observed power for n number of samples.

$$\overline{y(\iota)} = \frac{1}{n}\sum_{i=1}^{n} y_a(i) \tag{3.68}$$

3.7 Experiment and Results

3.7.1 Parameters Estimation of Weibull Distribution for Simulated Data

To evaluate the performance of the six aforementioned methods in parameters estimation of Weibull distribution for the wind power application, various sets of data with varying scale parameters, shape parameter, and sample size are randomly generated. Tables 3.1 to 3.3 present the average RMSE, chi-square, and K-S test between actual and estimated Weibull parameters for the six methods after replicating 100 times with sample size of 50, 500, and 10,000 each time. The critical values of K-S test at the 5% significance level at the given sample size are 0.1923, 0.0608, and 0.0136 respectively. With respect to the maximum error in CDF, neither of six estimation methods exceeds the corresponding critical value. This implies that shape and scale parameters estimated by these six methods are very close to the true parameter of the Weibull distribution. In other words, if two individual samples of data are randomly generated; one by true Weibull parameters and the other by estimated parameters, it can be deduced that both samples come from the same distribution. Similarly, according to critical value in chi-square error at the 5 % significance level with *df* degree of freedom, it can be concluded that sample data generated by true Weibull parameters and estimated parameters follow the same distribution. It should be mentioned that the critical value of chi-square error is extracted from the chi-square table presented in Appendix A. Tables 3.1 to 3.3 demonstrate that both K-S test and X^2 become smaller when the number of data increases. The similar trend is identified in RMSE value, irrespective of numerical methods. It can be also observed that RMSE values decrease while the scale parameter increases, irrespective of estimation methods, sample size, and shape parameter. Three other types of error, R-squared, percentage error of shape and scale parameters, for the same data number of 50, 500, and 10,000 are given in Tables 3.4 to 3.6. The results confirm the increment in R-squared error and a corresponding decrease in percentage error of shape and scale parameters with the size of samples. Note that all the results in Tables 3.1 to 3.6 are averaged over 100 replications.

Table 3.1: Mean of RMSE in PDF, mean of chi-square error, and mean of max-error in CDF between Weibull function and generated data after 100 replications with 50 random variables each time (Q_{95} = 0.1923), ($X^2|_{\alpha=0.05}^{df=4} = 9.48$)

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	RMSE	KS	X^2	RMSE	KS	X^2	RMSE	KS	X^2
1.5	2	0.0417	0.080	2.937	0.0353	0.0763	2.647	0.0350	0.0762	2.656
	5	0.0147	0.091	2.365	0.0150	0.0838	2.827	0.0150	0.0836	2.816
	10	0.0106	0.088	2.685	0.0089	0.0835	2.013	0.0089	0.0835	2.005
	15	0.0052	0.092	3.841	0.0062	0.0791	3.614	0.0062	0.0789	3.618
2	2	0.0350	0.1005	4.006	0.0312	0.0918	3.624	0.0305	0.0918	3.634
	5	0.0234	0.1033	3.470	0.0162	0.0817	2.715	0.0162	0.0814	2.716
	10	0.0073	0.0902	2.716	0.0080	0.0783	2.667	0.0079	0.0783	2.664
	15	0.0041	0.1023	4.061	0.0053	0.0792	3.618	0.0053	0.0793	3.627
2.5	2	0.0469	0.1091	3.1088	0.0482	0.0862	3.711	0.0479	0.0861	3.710
	5	0.0218	0.1005	4.5107	0.0200	0.0901	4.149	0.0197	0.0906	4.164
	10	0.0080	0.0956	4.7265	0.0081	0.0834	4.483	0.0081	0.0834	4.493
	15	0.0064	0.0949	3.5833	0.0062	0.0778	3.113	0.0062	0.0777	3.106
3	2	0.0608	0.1064	5.808	0.0400	0.0842	4.707	0.0394	0.0846	4.737
	5	0.0261	0.1141	5.559	0.0277	0.0875	4.702	0.0275	0.0875	4.724
	10	0.0086	0.0996	3.065	0.0082	0.0800	2.926	0.0082	0.0796	2.930
	15	0.0072	0.1136	7.723	0.0050	0.0877	6.014	0.0051	0.0876	6.032
True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.0348	0.0738	2.687	0.0372	0.0799	2.731	0.0458	0.0778	2.555
	5	0.0172	0.0832	2.661	0.0144	0.0875	3.317	0.0144	0.0815	2.181
	10	0.0087	0.0841	1.982	0.0090	0.0872	2.178	0.0096	0.0820	2.225
	15	0.0064	0.0780	3.417	0.0063	0.0828	3.839	0.0067	0.0862	2.958
2	2	0.0299	0.0946	3.833	0.0322	0.0931	3.721	0.0527	0.0915	3.295
	5	0.0170	0.0822	3.539	0.0154	0.0826	2.758	0.0173	0.0788	2.561
	10	0.0082	0.0798	2.667	0.0079	0.0788	3.140	0.0080	0.0794	2.478
	15	0.0053	0.0793	3.660	0.0055	0.0793	3.764	0.0049	0.0873	3.785
2.5	2	0.0531	0.0880	3.905	0.0471	0.0856	3.672	0.0461	0.0885	3.274
	5	0.0201	0.0903	3.220	0.0181	0.0926	4.425	0.0236	0.0889	4.260
	10	0.0080	0.0854	4.519	0.0081	0.0831	4.540	0.0098	0.0825	3.835
	15	0.0064	0.0797	3.153	0.0060	0.0768	3.124	0.0062	0.0867	3.422
3	2	0.0420	0.0848	4.830	0.0300	0.0872	5.162	0.0536	0.0805	4.182
	5	0.0286	0.0868	5.111	0.0239	0.0947	5.533	0.0346	0.0929	5.218
	10	0.0089	0.0809	2.965	0.0079	0.0802	3.136	0.0105	0.0899	2.878
	15	0.0050	0.0882	6.033	0.0051	0.0891	6.138	0.0060	0.0872	5.535

Table 3.2: Mean of RMSE in PDF, mean of chi-square error, and mean of max-error in CDF between Weibull function and generated data after 100 replications with 500 random variables each time (Q_{95} = 0.0608), ($X^2|_{\alpha=0.05}^{df=7} = 14.06$)

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	RMSE	KS	X^2	RMSE	KS	X^2	RMSE	KS	X^2
1.5	2	0.0117	0.0343	6.916	0.0107	0.0288	6.172	0.0107	0.0287	6.242
	5	0.0046	0.0321	7.056	0.0054	0.0274	6.950	0.0052	0.0272	6.879
	10	0.0026	0.0327	6.486	0.0024	0.0309	5.913	0.0023	0.0306	5.889
	15	0.0016	0.0321	6.877	0.0015	0.0285	6.767	0.0014	0.0283	6.712
2	2	0.0172	0.0351	5.579	0.0111	0.0284	3.821	0.0107	0.0280	3.732
	5	0.0054	0.0361	6.343	0.0045	0.0279	5.453	0.0044	0.0276	5.390
	10	0.0034	0.0433	7.114	0.0025	0.0293	6.356	0.0025	0.0292	6.264
	15	0.0017	0.0323	6.246	0.0014	0.0269	5.724	0.0014	0.0268	5.626
2.5	2	0.0212	0.0399	7.692	0.0116	0.0249	6.247	0.0114	0.0249	6.213
	5	0.0058	0.0381	7.237	0.0054	0.0267	5.648	0.0052	0.0268	5.610
	10	0.0039	0.0361	8.280	0.0026	0.0286	7.450	0.0026	0.0282	7.327
	15	0.0023	0.0391	7.992	0.0019	0.0272	6.757	0.0019	0.0272	6.679
3	2	0.0178	0.0368	8.347	0.0141	0.0239	6.045	0.0138	0.0238	5.993
	5	0.0090	0.0397	9.612	0.0052	0.0286	6.638	0.0049	0.0285	6.555
	10	0.0045	0.0435	8.467	0.0024	0.0256	6.117	0.0023	0.0260	6.052
	15	0.0029	0.0415	9.200	0.0019	0.0249	6.618	0.0019	0.0250	6.879
True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.0111	0.0273	6.017	0.0112	0.0313	6.456	0.0118	0.0257	6.081
	5	0.0045	0.0260	6.523	0.0044	0.0274	6.850	0.0053	0.0240	6.641
	10	0.0018	0.0297	5.978	0.0026	0.0306	6.124	0.0022	0.0276	6.022
	15	0.0014	0.0266	6.475	0.0014	0.0292	6.765	0.0016	0.0252	6.851
2	2	0.0104	0.0281	3.764	0.0113	0.0283	3.765	0.0092	0.0249	4.026
	5	0.0045	0.0274	5.322	0.0046	0.0280	5.5031	0.0049	0.0258	5.557
	10	0.0023	0.0292	6.108	0.0023	0.0299	6.217	0.0029	0.0254	5.597
	15	0.0014	0.0274	5.501	0.0014	0.0270	5.601	0.0014	0.0248	5.582
2.5	2	0.0114	0.0256	6.209	0.0106	0.0258	6.338	0.0130	0.0244	6.500
	5	0.0051	0.0282	5.623	0.0046	0.0279	5.727	0.0059	0.0274	6.000
	10	0.0025	0.0284	7.242	0.0024	0.0283	7.384	0.0026	0.0275	7.6014
	15	0.0019	0.0279	6.717	0.0018	0.0284	6.688	0.0020	0.0254	7.220
3	2	0.0138	0.0243	6.003	0.0168	0.0304	7.447	0.0147	0.0241	6.165
	5	0.0050	0.0288	6.552	0.0064	0.0322	8.066	0.0061	0.0259	6.831
	10	0.0024	0.0263	6.066	0.0034	0.0328	7.509	0.0025	0.0243	6.465
	15	0.0021	0.0260	6.579	0.0027	0.0310	8.132	0.0018	0.0233	6.521

Table 3.3: Mean of RMSE in PDF, mean of chi-square error, and mean of max-error in CDF between Weibull function and generated data after 100 replications with 10000 random variables each time (Q_{95}= 0.0136), ($X^2|_{\alpha=0.05}^{df=8} = 15.50$)

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	RMSE	KS	X^2	RMSE	KS	X^2	RMSE	KS	X^2
1.5	2	0.0041	0.0090	8.812	0.0048	0.0088	8.071	0.0044	0.0084	7.281
	5	0.0009	0.0081	7.307	0.0021	0.0087	8.941	0.0019	0.0082	8.091
	10	0.0007	0.0082	10.412	0.0010	0.0096	9.820	0.0009	0.0091	8.978
	15	0.0004	0.0095	9.057	0.0006	0.0101	10.835	0.0005	0.0096	9.965
2	2	0.0037	0.0109	12.865	0.0053	0.0082	11.063	0.0045	0.0076	9.681
	5	0.0017	0.0106	12.151	0.0015	0.0074	9.769	0.0012	0.0068	8.486
	10	0.0008	0.0098	11.709	0.0009	0.0084	11.477	0.0008	0.0077	9.902
	15	0.0006	0.0084	7.332	0.0004	0.0080	9.948	0.0004	0.0073	8.427
2.5	2	0.0054	0.0101	9.509	0.0039	0.0072	7.689	0.0033	0.0064	6.737
	5	0.0029	0.0116	16.409	0.0018	0.0070	8.971	0.0015	0.0064	7.679
	10	0.0010	0.0110	12.243	0.0008	0.0065	8.763	0.0006	0.0059	7.446
	15	0.0006	0.0097	10.506	0.0006	0.0067	7.110	0.0005	0.0059	5.933
3	2	0.0066	0.102	14.282	0.0034	0.0062	8.253	0.0031	0.0057	7.752
	5	0.0034	0.0122	18.455	0.0018	0.0060	7.340	0.0016	0.0059	6.994
	10	0.0023	0.0169	28.319	0.0007	0.0071	8.021	0.0007	0.0069	7.427
	15	0.0011	0.0118	23.957	0.0003	0.0059	7.889	0.0003	0.0055	7.317

True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.0026	0.0070	4.971	0.0029	0.0072	5.053	0.0028	0.0064	4.716
	5	0.0006	0.0062	5.559	0.0007	0.0062	5.393	0.0008	0.0058	5.475
	10	0.0004	0.0064	5.982	0.0004	0.0064	5.999	0.0004	0.0058	6.206
	15	0.0003	0.0072	6.939	0.0003	0.0071	6.943	0.0004	0.0063	6.761
2	2	0.0022	0.0066	7.753	0.0033	0.0070	8.405	0.0026	0.0056	7.905
	5	0.0008	0.0063	6.927	0.0009	0.0062	7.323	0.0008	0.0055	6.897
	10	0.0006	0.0069	7.285	0.0006	0.0068	8.367	0.0005	0.0062	7.511
	15	0.0004	0.0061	5.921	0.0003	0.0063	6.752	0.0004	0.0055	6.116
2.5	2	0.0032	0.0056	6.407	0.0039	0.0057	7.321	0.0035	0.0051	6.725
	5	0.0013	0.0065	7.040	0.0014	0.0070	7.712	0.0014	0.0055	7.797
	10	0.0005	0.0063	6.717	0.0005	0.0069	7.270	0.0006	0.0057	6.918
	15	0.0004	0.0051	5.439	0.0004	0.0056	6.283	0.0004	0.0050	5.808
3	2	0.0032	0.0057	7.755	0.0162	0.0133	39.188	0.0035	0.0053	7.987
	5	0.0015	0.0059	6.968	0.0064	0.0144	38.925	0.0018	0.0056	6.880
	10	0.0008	0.0069	7.342	0.0034	0.0133	36.003	0.008	0.0059	7.784
	15	0.0003	0.0056	7.249	0.0021	0.0131	37.890	0.0003	0.0050	7.575

Table 3.4: Mean of R-squared and percent errors (p.e) of Weibull parameters after 100 replications with 50 random variables each time

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	R^2	p.e k	p.e c	R^2	p.e k	p.e c	R^2	p.e k	p.e c
1.5	2	0.7576	15.07	7.17	0.799	12.89	5.83	0.8032	12.73	5.82
	5	0.7863	9.49	6.40	0.7865	8.25	7.75	0.7853	8.27	7.76
	10	0.7069	17.68	9.89	0.8003	14.50	9.19	0.8014	14.45	9.19
	15	0.8021	12.19	8.59	0.6925	14.97	9.17	0.6974	14.76	9.18
2	2	0.8667	10.56	5.01	0.8745	9.41	3.91	0.8794	9.08	3.91
	5	0.6990	18.48	8.34	0.8451	10.60	7.02	0.8476	10.51	7.03
	10	0.8376	8.94	5.57	0.7862	8.51	6.31	0.7893	8.26	6.31
	15	0.8908	6.65	5.61	0.8339	10.51	6.29	0.8375	10.36	6.29
2.5	2	0.8313	6.28	7.50	0.8081	9.76	5.67	0.8124	9.55	5.67
	5	0.7854	12.92	6.27	0.8231	10.68	5.78	0.8289	10.42	5.78
	10	0.8730	8.63	4.64	0.8879	7.19	5.54	0.8876	7.26	5.57
	15	0.8504	7.58	6.69	0.8310	6.49	7.06	0.8331	6.27	7.06
3	2	0.8145	14.32	4.18	0.9925	9.21	2.49	0.9257	9.02	2.50
	5	0.7647	11.02	5.97	0.7396	12.49	6.38	0.7460	12.71	6.39
	10	0.9081	7.77	3.34	0.9121	5.92	4.22	0.9137	5.76	4.23
	15	0.8723	11.07	4.19	0.9329	7.67	3.49	0.9325	7.80	3.48
True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.8128	12.10	5.76	0.7924	13.87	5.87	0.7076	16.32	7.56
	5	0.6994	9.49	8.23	0.8189	7.67	7.56	0.7957	8.96	6.95
	10	0.8132	12.86	9.40	0.7973	15.34	9.04	0.7427	13.89	11.20
	15	0.6837	15.24	9.24	0.6879	15.57	9.17	0.6574	18.85	6.81
2	2	0.8862	8.97	3.72	0.8669	10.06	3.92	0.7039	16.04	5.80
	5	0.8314	11.34	7.10	*0.8587*	*9.62*	7.02	0.8626	8.37	6.89
	10	0.7680	9.10	6.38	0.7983	8.13	6.28	0.7887	9.68	5.39
	15	0.8355	10.25	6.28	0.8188	11.51	6.29	0.8507	8.81	5.54
2.5	2	0.7762	11.89	5.64	0.8213	9.00	5.66	0.8288	6.26	6.48
	5	0.8241	11.09	5.78	0.8637	9.21	5.81	0.7509	13.37	6.40
	10	0.8915	7.40	5.41	0.8881	7.27	5.25	0.8377	9.09	5.87
	15	0.8253	6.89	7.07	0.8378	5.61	7.07	0.8274	7.73	6.42
3	2	0.9127	9.71	2.46	0.9570	5.29	2.64	0.8604	12.51	3.11
	5	0.7086	13.41	6.42	0.8291	8.92	6.52	0.6251	16.16	6.80
	10	0.8962	7.14	4.27	0.9253	5.25	4.32	0.8620	9.19	4.24
	15	0.9284	7.36	3.57	*0.9335*	7.52	3.48	0.9416	10.18	3.32

Table 3.5: Mean of R-squared and percent errors (p.e) of Weibull parameters after 100 replications with 500 random variables each time

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	R^2	p.e k	p.e c	R^2	p.e k	p.e c	R^2	p.e k	p.e c
1.5	2	0.9853	2.94	2.82	0.9883	2.67	2.48	0.9884	2.61	2.47
	5	0.9837	2.85	3.23	0.9787	4.39	2.33	0.9796	4.21	2.33
	10	0.9787	4.38	2.18	0.9813	3.57	2.19	0.9819	3.51	2.19
	15	0.9805	2.98	3.49	0.9822	2.68	2.69	0.9827	2.56	2.68
2	2	0.9712	5.21	2.05	0.9872	3.41	1.48	0.9883	3.26	1.47
	5	0.9805	3.49	1.82	0.9842	2.48	1.79	0.9847	2.40	1.79
	10	0.9748	3.85	3.39	0.9848	2.47	2.44	0.9855	2.26	2.44
	15	0.9845	2.75	2.38	0.9906	2.25	1.93	0.9910	2.17	1.93
2.5	2	0.9658	4.451	2.610	0.9890	2.100	1.506	0.9893	2.082	1.505
	5	0.9855	2.898	1.843	0.9870	2.514	1.658	0.9880	2.408	1.660
	10	0.9763	4.882	1.941	0.9894	2.695	1.456	0.9897	2.588	1.457
	15	0.9749	3.107	2.356	0.9852	1.878	2.105	0.9859	1.692	2.110
3	2	0.9826	3.855	1.305	0.9909	2.772	1.116	0.9915	2.745	1.108
	5	0.9721	5.513	1.309	0.9903	2.496	1.028	0.9913	2.171	1.032
	10	0.9716	4.490	1.785	0.9928	1.599	1.260	0.9934	1.410	1.256
	15	0.9747	4.596	1.617	0.9891	3.021	1.162	0.9892	3.092	1.154
True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.9882	3.18	2.42	0.9860	2.76	2.41	0.9868	3.79	2.34
	5	0.9839	3.15	2.45	0.9853	3.19	2.30	0.9784	4.41	2.14
	10	0.9980	2.41	2.12	0.9798	4.08	2.17	0.9851	3.06	2.40
	15	0.9842	2.45	2.78	0.9837	2.90	2.63	0.9802	3.57	2.74
2	2	0.9892	3.04	1.50	0.9875	3.55	1.47	0.9929	2.60	1.52
	5	0.9848	2.44	1.79	0.9845	2.49	1.79	0.9820	2.55	2.06
	10	0.9862	2.20	2.44	0.9864	1.88	2.44	0.9813	3.36	2.39
	15	0.9906	2.18	1.97	0.9910	2.34	1.92	0.9906	2.53	1.67
2.5	2	0.9891	2.111	1.516	*0.9898*	1.795	1.513	0.9903	2.687	1.403
	5	0.9888	2.494	1.667	0.9908	1.839	1.673	0.9847	3.105	1.622
	10	0.9898	2.572	1.460	*0.9905*	2.380	1.468	0.9912	2.409	1.436
	15	0.9861	1.866	2.104	0.9862	1.734	2.127	0.9849	2.222	2.127
3	2	0.9915	2.685	1.120	0.9870	3.979	1.051	0.9892	2.865	1.192
	5	0.9911	2.301	1.040	0.9885	3638	1.076	0.9880	3.248	1.104
	10	0.9928	1.669	1.272	0.9874	3.714	1.241	0.9932	1.712	1.289
	15	0.9880	3.451	1.141	0.9820	4.804	1.100	0.9899	2.629	1.156

Table 3.6: Mean of R-squared and percent errors (p.e) of Weibull parameters after 100 replications with 10,000 random variables each time

True Weibull parameters		Graphical method			Moment method			Empirical method		
k	c (m/s)	R^2	p.e k	p.e c	R^2	p.e k	p.e c	R^2	p.e k	p.e c
1.5	2	0.9981	1.300	0.961	0.9974	1.8588	0.645	0.9978	1.661	0.649
	5	0.9993	0.7322	0.574	0.9974	1.861	0.675	0.9978	1.663	0.658
	10	0.9982	1.274	0.558	0.9972	2.008	0.499	0.9976	1.810	0.490
	15	0.9987	0.940	0.649	0.9979	1.664	0.641	0.9982	1.505	0.650
2	2	0.9987	1.102	0.468	0.9977	1.887	0.412	0.9983	1.555	0.411
	5	0.9981	1.138	0.611	0.9987	1.261	0.345	0.9991	0.954	0.345
	10	0.9983	1.126	0.574	0.9979	1.505	0.547	0.9984	1.177	0.547
	15	0.9980	1.070	0.843	0.9987	1.076	0.502	0.9990	0.911	0.523
2.5	2	0.9980	1.226	0.672	0.9990	0.957	0.401	0.9993	0.701	0.399
	5	0.9961	1.453	0.974	0.9987	1.186	0.377	0.9990	0.924	0.380
	10	0.9981	1.272	0.515	0.9987	1.240	0.322	0.9992	0.828	0.325
	15	0.9983	1.193	0.463	0.9986	1.204	0.442	0.9990	0.942	0.440
3	2	0.9977	1.669	0.299	0.9994	0.737	0.200	0.9995	0.693	0.204
	5	0.9968	1.756	0.525	0.9991	0.764	0.404	0.9993	0.601	0.391
	10	0.9929	2.796	0.722	0.9993	0.766	0.251	0.9992	0.827	0.237
	15	0.9956	2.286	0.216	0.9996	0.523	0.194	0.9997	0.512	0.181
True Weibull parameter		Maximum likelihood			Energy pattern factor			GSO algorithm		
1.5	2	0.9992	0.581	0.717	0.9991	0.750	0.695	0.9992	0.832	0.680
	5	0.9996	0.255	0.515	0.9996	0.406	0.524	0.9994	0.413	0.557
	10	0.9995	0.542	0.457	0.9994	0.708	0.461	0.9995	0.780	0.451
	15	0.9991	0.604	0.695	0.9988	0.626	0.733	0.9988	0.761	0.776
2	2	0.9995	0.477	0.398	0.9990	0.993	0.413	0.9993	0.539	0.517
	5	0.9995	0.561	0.344	0.9995	0.606	0.345	0.9995	0.494	0.385
	10	0.9991	0.713	0.548	0.9988	0.801	0.548	0.9992	0.585	0.543
	15	0.9990	0.740	0.534	0.9992	0.735	0.526	0.9991	0.702	0.528
2.5	2	0.9993	0.626	0.398	0.9990	0.886	0.394	0.9993	0.824	0.371
	5	0.9992	0.673	0.986	0.9991	0.843	0.389	0.9990	0.926	0.369
	10	0.9995	0.578	0.321	0.9994	0.646	0.331	0.9992	0.804	0.354
	15	0.9993	0.588	0.435	0.9992	0.567	0.436	0.9993	0.653	0.443
3	2	0.9995	0.722	0.203	0.9891	4.451	0.256	0.9993	0.771	0.211
	5	0.9993	0.584	0.388	0.9896	4.305	0.345	0.9993	0.733	0.343
	10	0.9992	0.852	0.243	0.9880	4.706	0.223	0.9990	0.848	0.254
	15	0.9997	0.353	0.180	0.9897	4.349	0.193	0.9996	0.415	0.204

For the sake of clarity and conciseness, all the results are summarized in Figures 3.9 to 3.10. The figures present how frequent each numerical method outperforms others in statistical tests for a particular sample size, shape and scale parameters. In this regard, when the data number is 50, EPFM overall achieves better results than the other numerical methods in terms of RMSE and p.e k while EM is the worst. With respect to K-S test and chi-square error, unlike GM and EPFM respectively, GSO algorithm shows the best performance. GM followed by MLM and EPFM demonstrate the satisfactory results in terms of R-squared while in terms of p.e c, GSO followed by MLM and GM perform more accurately compared to other numerical methods. Figure 3.7 also verifies that in the determination of the best numerical method, no firm conclusion can be reached.

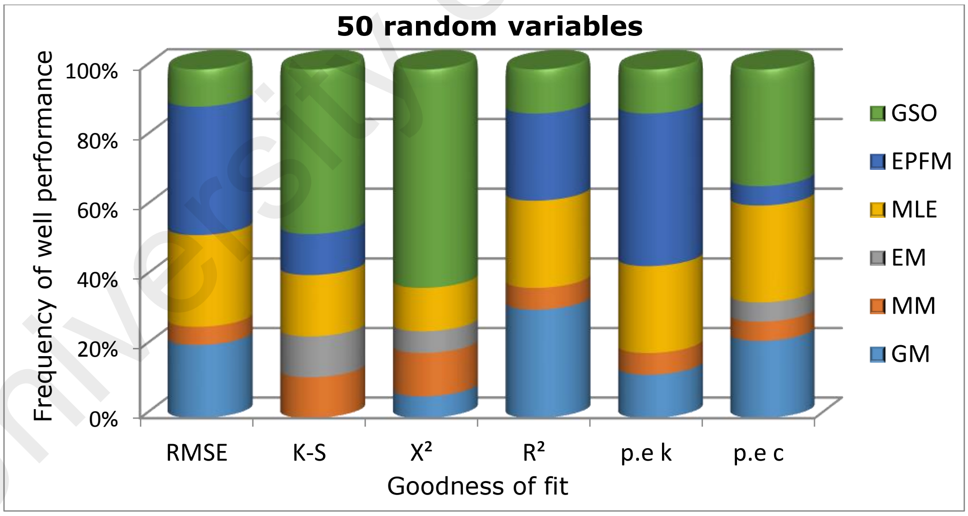


Figure 3.9: The best performance of the parameter estimation methods with 50 random samples

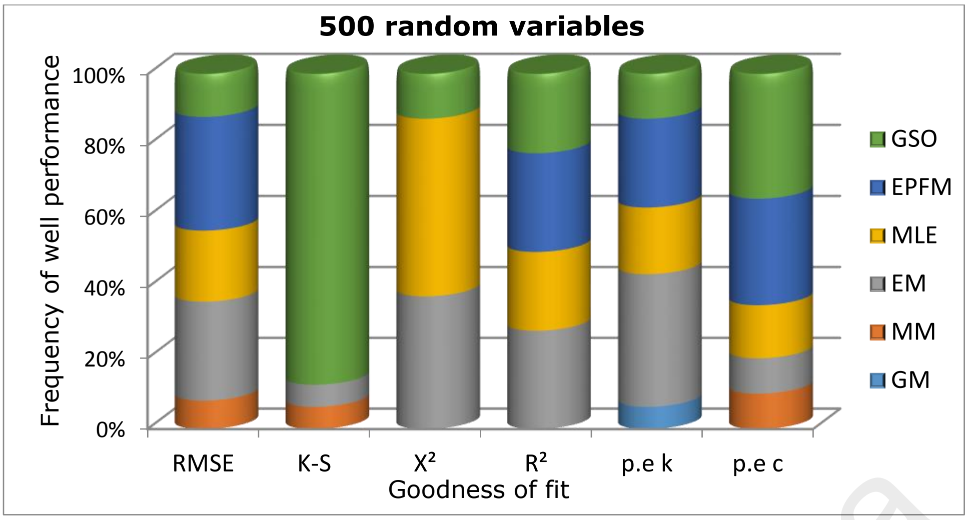


Figure 3.10: The best performance of the parameter estimation methods with 500 random samples

As the number of data increases to 500, the obtained results by numerical methods show the different pattern with the one at the sample size of 50. Figure 3.10 proves that GM and MM overall have the worst performance while EM performs more accurately as compared to when data number is 50. The similar pattern is observed in obtained results by EPFM except in K-S test. With respect to K-S test, MLM performs less accurately, but overall its performance is improved. GSO presents highly satisfactory results in terms of chi-square, however, its performance is degraded in terms of K-S test and p.e c as compared to Figure 3.9.

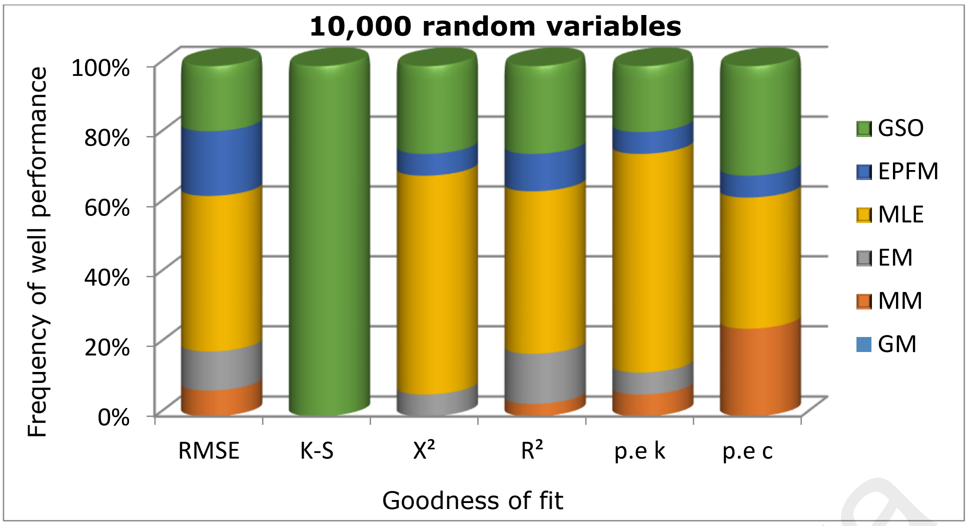


Figure 3.11: The best performance of the parameter estimation methods with 10,000 random samples

For a large number of data, all numerical methods yield better results as compared to when the sample size is 500. As shown in Figure 3.11, MLM followed by GSO is far superior to other estimation methods. MM, EMPF, and EM illustrate the same performance, while the GM is the worst.

Figures 3.12 and 3.13 depict the impact of shape and scale parameters on Weibull distribution. The mean of data is described by the scale parameter while the flatness of Weibull distribution is represented by the shape parameter. In other words, the scale parameter controls the Abscissa of Weibull distribution and the shape parameter affects narrowness and the peak value of the Weibull distribution.

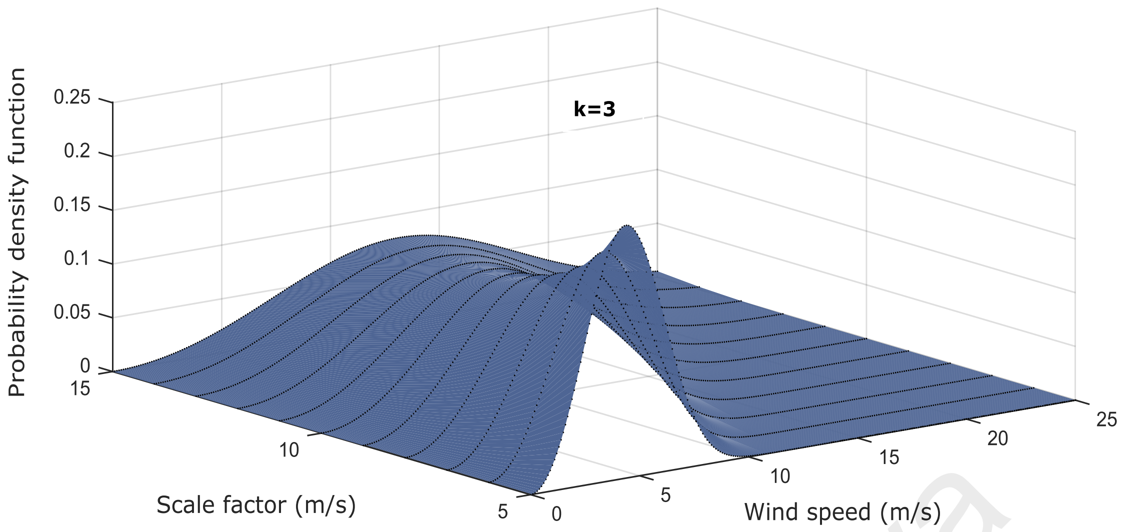


Figure 3.12: Weibull distribution for the shape factor of 3 and scale factor between 5 and 15

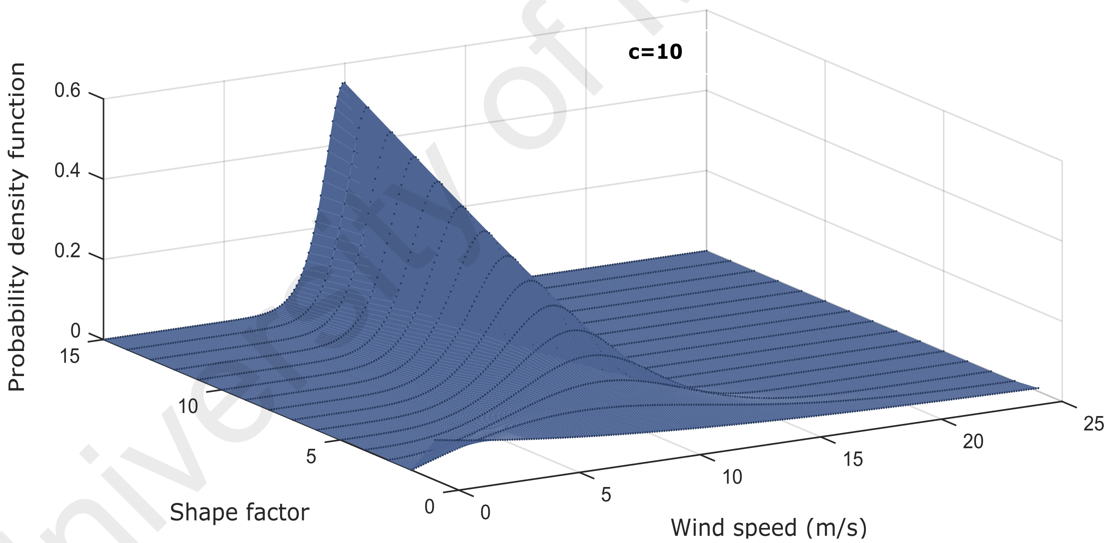


Figure 3.13: Weibull distribution for the scale factor of 10 and shape factor between 1 and 15

3.7.2 Parameters Estimation of Weibull Distribution for Actual Data

The wind farm selected in this study, named Khaf, is a wind farm situated in onshore of Razavi Khorasan province of Iran, 250 km southwest of Mashhad. It is situated at 34°.34' N latitude and 60°.8' E longitude in height of 966 m above sea level. Khaf has a great potential of wind energy, an average capacity factor of over 45%, while according to International Energy Agency (IEA) most wind power plants operate at a capacity factor of 25–40%. All wind data of this station are measured with the use of either a combination of a wind vane and a cup anemometer or a sonic anemometer. The wind turbine considered is the regulated-pitch type 1.5 MW WD77 with the hub height of 60 m, the cut-in speed rate of 3 m/s, the rated speed of 11 m/s, and the cut-out speed rate of 25 m/s. Three hundred thirty days of data were acquired with a total of 94,271 5-min observations comprising wind speed, actual power, and wind direction. The data set was collected for four seasons starting from 09.04.2013 15:05:00 P.M to 04.03.2014 23:55:00 P.M.

Generally, the data collected from wind farms contains various noises and anomalies owing to reasons such as sensor errors, blade damage, maintenance issues, malfunction of the pitch control, low wind speed, inappropriate setting of the pitch angle, fluctuations in the turbine performance, and environment issues (dirt, ice, etc.). Thus, to build a robust, stable, and accurate model for a wind turbine, data filtering is required. As illustrated in Figure 3.14, wind speed data is divided into intervals as small as 0.2 m/s, e.g., the intervals in [4.8–5 m/s]. In the next step, the standard deviation and average (μ_y, σ_y) of the corresponding wind powers are calculated. Those located outside the boundary $[\mu_y - \sigma_y, \mu_y + \sigma_y]$ are filtered, shown as white circled data in Figure 3.14 (b). To improve the data quality, in the third step, wind powers are equally split into bins of ten where each bin represents the number of occurrence of wind power in that particular interval. Powers with lower probability than the given threshold are discarded. $\sum \rho_i > \rho_T$ where ρ_T is threshold probability, i is the number of bins, and q is the sorted probability. The pink dots and bins in Figure 3.14 (b) and (c) are the selected data with high probabilities. The filtered power curve can then be obtained by sliding the wind power over all the wind speed intervals, as shown in Figure 3.14 (a).

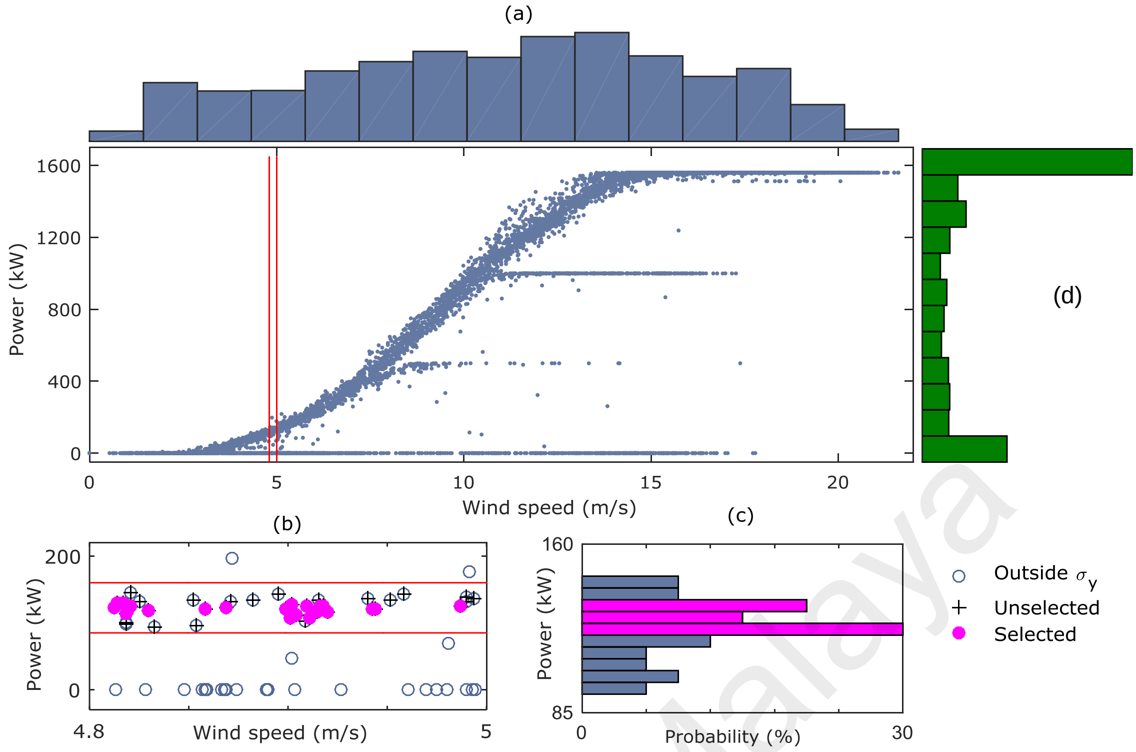


Figure 3.14: (a) 5-min average value of collected data in April, wind speed distribution, and frequency of the turbine power, (b and c) process of filtering out the outliers, (d) frequency of the output power

Table 3.7 lists the statistical characteristic of the collected data including 12 months in 2014. In this table, mean value shows the central value of the observed wind speed. The standard deviation describes how spread out the wind speeds is. The skewness is the extent to which the wind speed data are not symmetrical. If the value of the skewness is close to zero it indicates that data is more symmetrical, though the low value of skewness alone does not imply that data follows the normal distribution. The tail of the distribution points to the right (right-skewed data) if the value of skewness is greater than 0. The left skewed data, by contrast, produces the negative skewness value. Kurtosis describes how the tail and peak of a sample data (distribution) differ from a normal distribution. The more kurtosis value approaches to zero, the more data follow a normal distribution. A positive value of kurtosis shows the sample data has a sharper peak and heavier tails whereas the kurtosis negative value indicates the flatter peak and lighter tails than a normal distribution.

Table 3.7: Descriptive statistics for monthly, seasonal, and yearly wind data observed at Khaf

Period	Data no.	Mean (m/s)	Standard deviation (m/s)	Skewness	Kurtosis	Q_{95}
January	8928	5.50	3.87	1.221	4.471	0.0144
February	8061	6.65	5.64	1.108	3.316	0.0151
March	1152	6.09	4.48	0.517	1.721	0.0401
April	5637	10.52	6.44	0.943	3.52	0.0181
May	8928	12.46	4.95	-0.285	2.560	0.0144
June	8640	15.43	4.39	-0.889	3.345	0.0146
July	8908	12.82	4.25	-0.059	2.536	0.0144
August	8925	13.99	4.17	-0.456	3.092	0.0144
September	8640	9.19	5.07	0.061	2.021	0.0146
October	8903	8.99	6.06	0.291	2.008	0.0144
November	8640	5.72	4.67	1.167	3.774	0.0146
December	8908	5.85	4.73	1.059	3.285	0.0144
October-December	26451	6.91	5.4	0.821	2.709	0.0084
April-June	23205	13.31	5.06	-0.453	2.473	0.0089
July-September	26473	12.08	4.92	-0.311	2.518	0.0084
January-March	18141	6.05	4.81	1.261	4.149	0.0101
Yearly	94270	9.89	5.94	0.12	1.88	0.0044

In this section, to evaluate the wind energy resources in Khaf site four different probability distributions, namely, Weibull, gamma, Rayleigh, and log-normal, are applied to fit the observed wind speed frequency. To estimate the parameters of mentioned probability distributions three of aforementioned numerical methods in 3.4.2 (i.e MM, MLM, and GSO) are employed. To draw a fair comparison between numerical methods and probability distributions, R-squared, RMSE, K-S test, and chi-square are used as the goodness-of-fit, as listed in Tables 3.8 and 3.9. Based on the obtained results, it can be deduced that GSO and MLM overall outperform MM. In addition, a larger value of R-squared and lower value of RMSE for Weibull distribution indicate the superiority of Weibull as compared to other probability distributions. In other words, it shows that historical wind speed data in winter follows Weibull distributions more than others. According to Table 3.8, both K-S and X^2 statistics obtained by MM-Rayleigh are greater than the critical values at the 5% significance level, 0.0101 and 15.50 respectively. Therefore, MM estimation method did not pass the two statistical tests suggesting that observed wind data in winter does not follow this probability distribution. As given in Table 3.9, relatively similar pattern of results is obtained based on fall data. Based on the results, although all the numerical methods pass the statistical tests, Weibull distribution exhibits considerably better fit to the observed wind speed irrespective of parameter estimation methods. The estimated parameters of probability distributions using GSO, MLM, and MM based on spring, summer, and yearly wind speed data are presented in Table 3.10.

Table 3.8: Goodness-of-fit of the distribution models based on MM, MLE, DE in winter, ($\mathbf{Q_{95}}$=0.0101), ($\boldsymbol{X^2}|_{\boldsymbol{\alpha=0.05}}^{\boldsymbol{df=8}} = \mathbf{15.50}$)

Models	Goodness of fit	Estimator		
		MM	MLM	GSO
Weibull	R^2	0.9264	0.9422	0.9431
	RMSE	0.0099	0.0083	0.0086
	K-S	0.0070	0.0065	0.0073
	X^2	13.06	12.88	12.62
Gamma	R^2	0.9102	0.9112	0.9207
	RMSE	0.0105	0.0120	0.0109
	K-S	0.0082	0.0071	0.0079
	X^2	14.03	13.94	13.27
Rayleigh	R^2	0.853	0.875	0.875
	RMSE	0.0219	0.0201	0.0215
	K-S	0.0113	0.0099	0.0097
	X^2	17.50	15.33	15.30
Log-normal	R^2	0.8851	0.8915	0.8980
	RMSE	0.0185	0.0188	0.0196
	K-S	0.0099	0.0084	0.0091
	X^2	16.01	15.27	14.91

Table 3.9: Goodness-of-fit of the distribution models based on MM, MLE, DE in fall ($\mathbf{Q_{95}}$=0.0084), ($X^2|_{\alpha=0.05}^{df=10} = \mathbf{18.307}$)

Models	Goodness of fit	Estimator		
		MM	MLM	GSO
Weibull*	R^2	0.9285	0.9319	0.9337
	RMSE	0.0073	0.0073	0.0077
	K-S	0.0010	0.0013	0.0011
	X^2	11.20	10.98	10.05
Gamma*	R^2	0.9011	0.9084	0.9123
	RMSE	0.0090	0.0097	0.0096
	K-S	0.0032	0.0038	0.0041
	X^2	12.90	13.11	12.09
Rayleigh*	R^2	0.8713	0.8817	0.9002
	RMSE	0.0142	0.0185	0.0189
	K-S	0.0069	0.0081	0.0059
	X^2	18.10	17.86	17.13
Log-normal*	R^2	0.8906	0.8988	0.9015
	RMSE	0.0145	0.0156	0.0124
	K-S	0.0074	0.0062	0.0056
	X^2	16.85	15.69	13.80

*The formulas are presented in Appendix B.

Table 3.10: Parameter estimation results of distribution models using MM, MLM, and DE estimation methods

Period	Estimation method	Weibull		Gamma		Rayleigh	Log-normal	
		c	k	c_1	k_1	c_2	μ	σ
Spring	MM	14.94	2.87	6.91	1.92	10.62	2.52	0.36
	MLE	14.87	2.91	4.58	2.90	10.07	2.47	0.55
	GSO	15.04	3.00	8.40	1.83	10.07	2.57	0.36
Summer	MM	13.59	2.65	6.00	2.01	9.64	2.41	0.39
	MLE	13.52	2.63	6.20	3.08	9.22	2.35	0.60
	GSO	13.91	2.79	7.80	1.79	9.22	2.47	0.039
Yearly	MM	11.11	1.74	3.56	2.77	7.89	2.13	0.55
	MLE	10.98	1.61	1.90	5.19	8.16	2.01	0.88
	GSO	11.55	1.62	1.60	7.85	8.16	2.16	0.68

Figures 3.15 to 3.20 present how four probability distribution functions whose parameters are estimated by GSO, MLM, and MM match the historical wind speed data observed in July and February. Additionally, the obtained CDF of each individual distribution function based on the corresponding estimated parameter are shown in the figures. Based on Figures 3.15 to 3.17, observed wind data in July deviates from

probability distribution functions, notably, from gamma, log-normal, and Rayleigh. According to Figures 3.18 to 3.20 Weibull distribution using GSO indicates better fitting to the historical wind data in February.

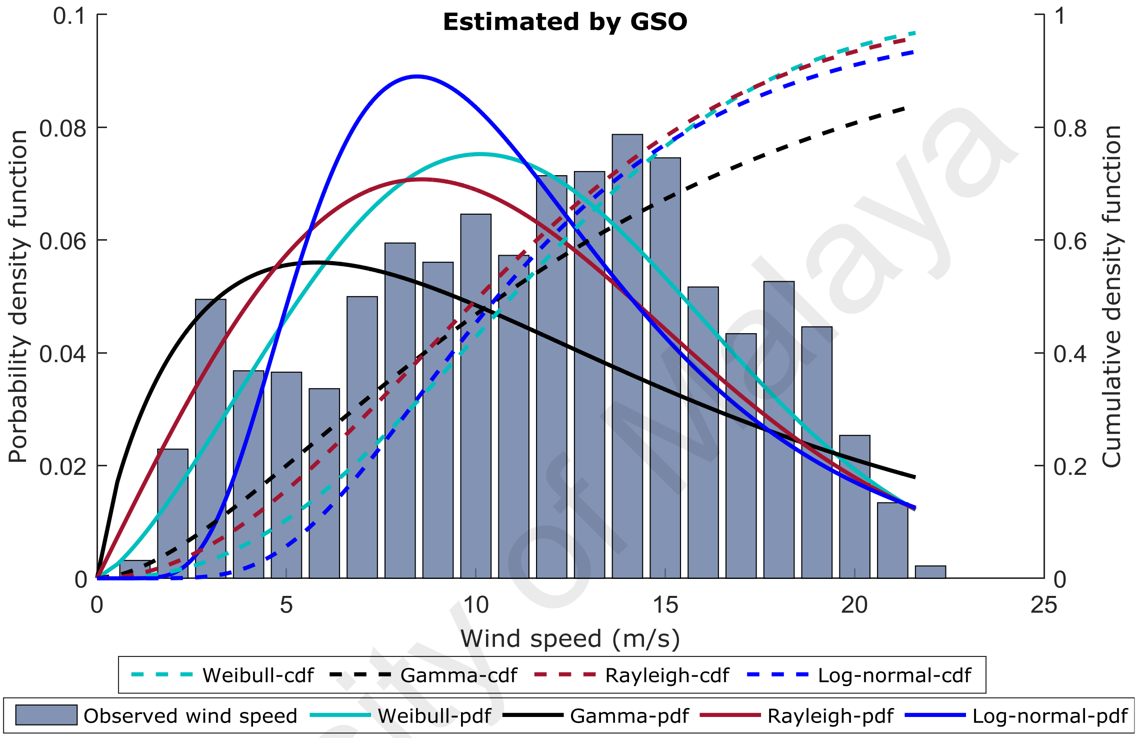


Figure 3.15: Observed wind speed in July and the estimated PDFs and CDFs of different distribution functions using GSO algorithm method

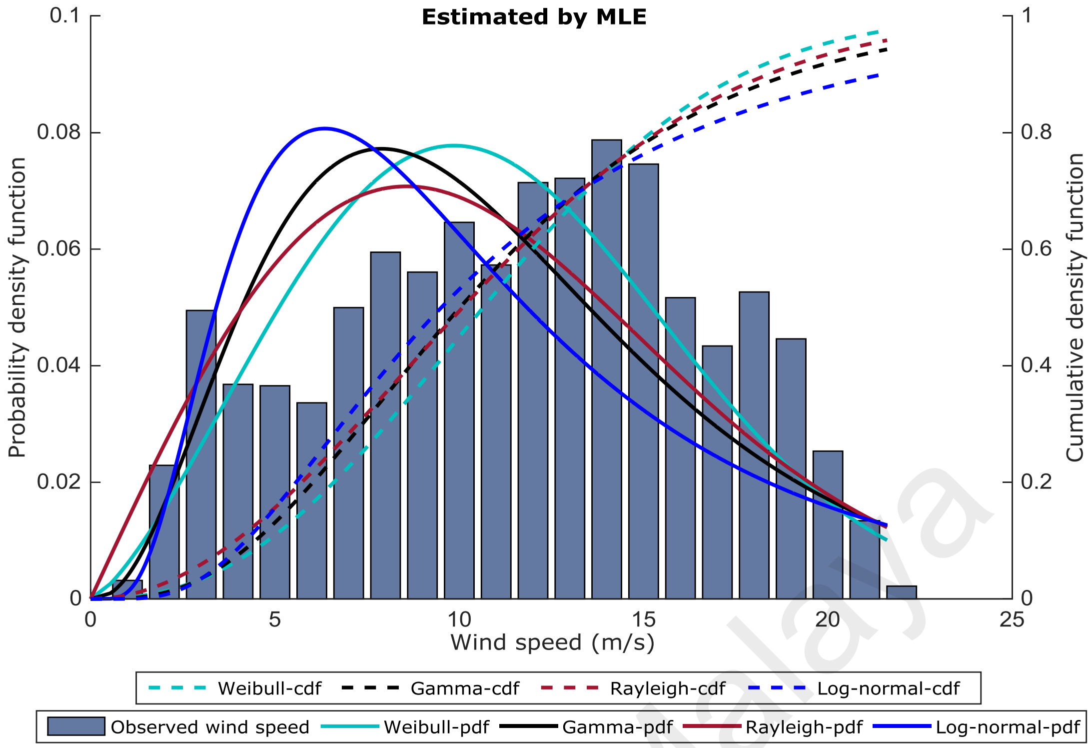


Figure 3.16: Observed wind speed in July and the estimated PDFs and CDFs of different distribution functions using MLM method

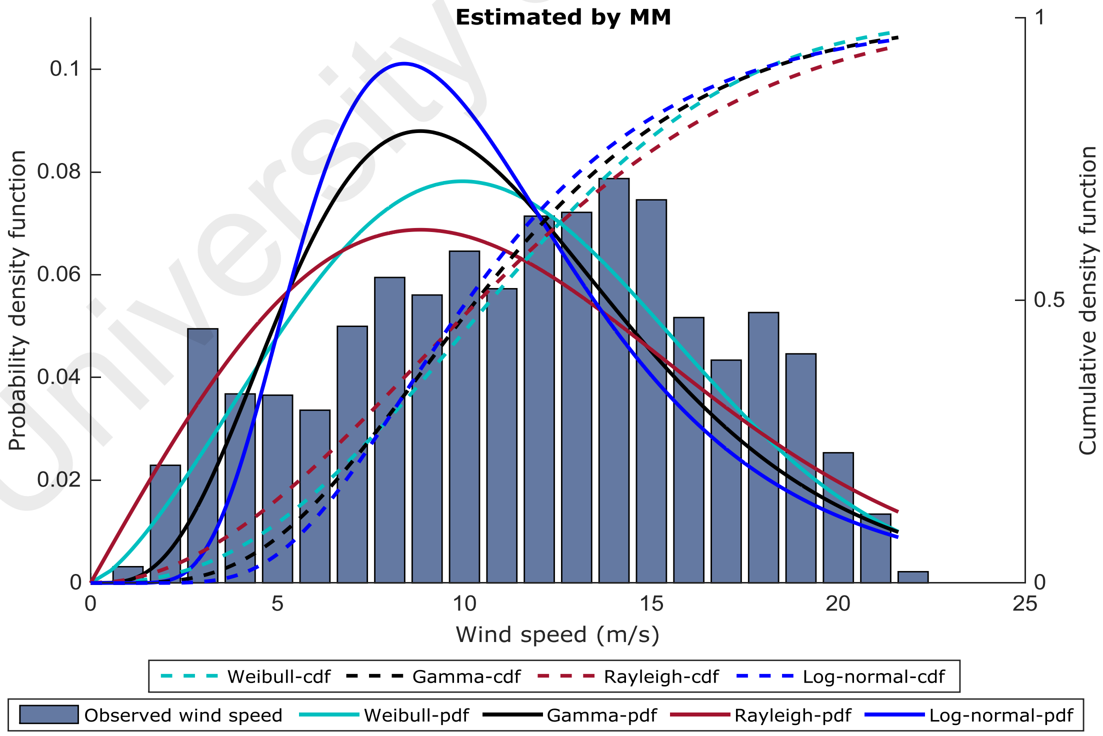


Figure 3.17: Observed wind speed in July and the estimated PDFs and CDFs of different distribution functions using MM method

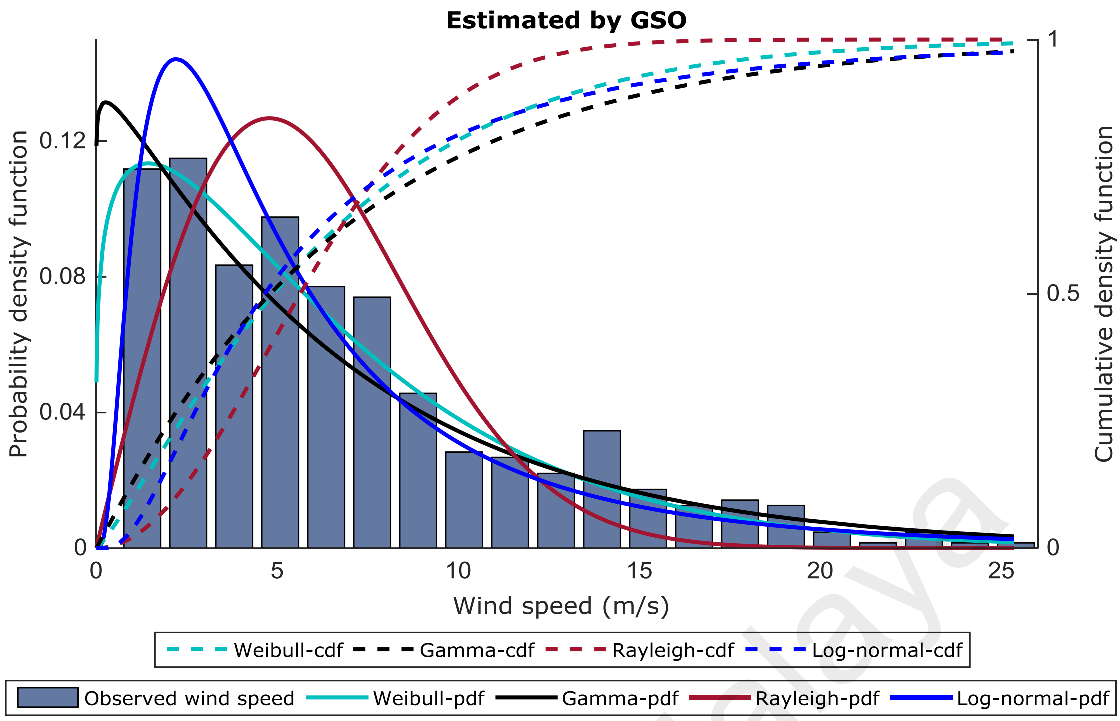


Figure 3.18: Observed wind speed in February and the estimated PDFs and CDFs of different distribution functions using GSO method

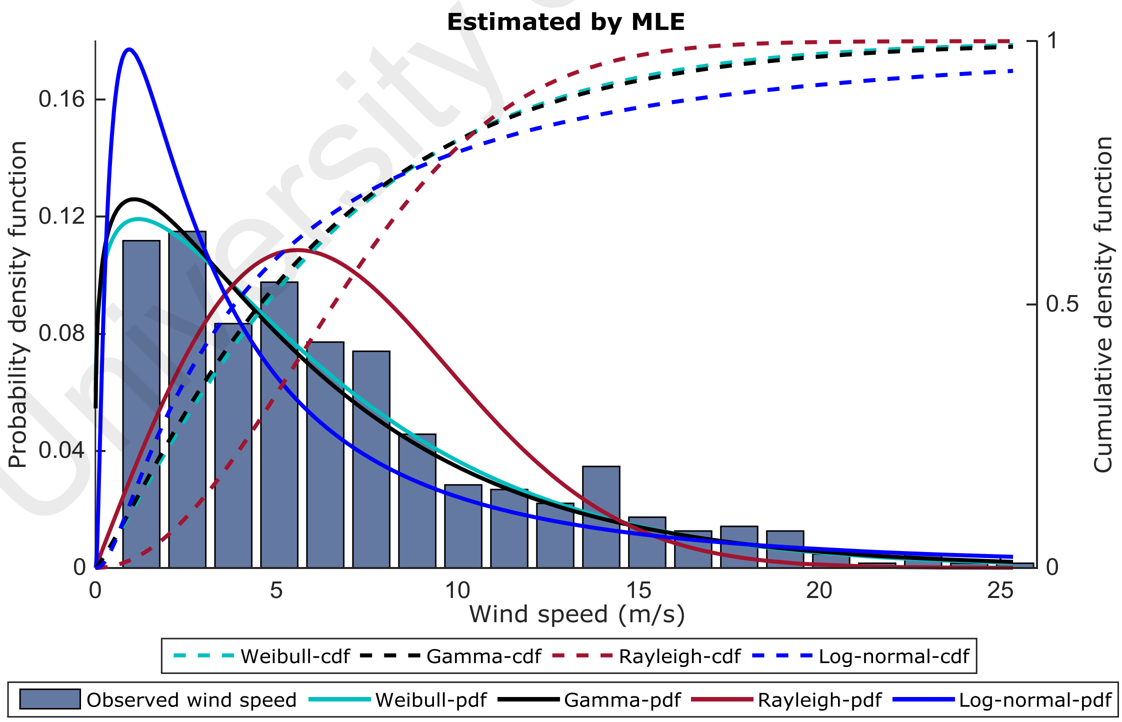


Figure 3.19: Observed wind speed in February and the estimated PDFs and CDFs of different distribution functions using MLM method

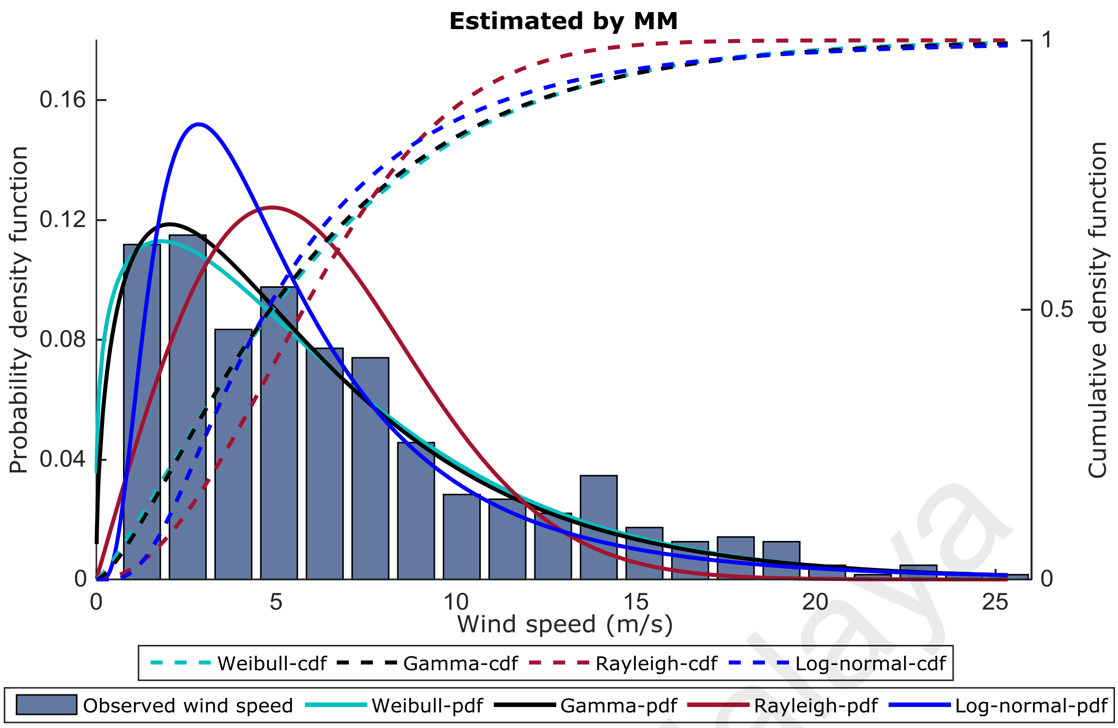


Figure 3.20: Observed wind speed in February and the estimated PDFs and CDFs of different distribution functions using MM method

The performance evaluation of the six numerical methods based on Weibull for 12-month as well as the corresponding estimated shape and scale parameters are listed in Tables 3.11 and 3.12. In the view of obtained criteria, GSO and MLM outperform other methods for most of the months except November, whereas the GM shows the worst performance. Based on obtained results from the tables, GM and EM for months May-July, as well as MM for month July could not pass the K-S statistical test. On one hand, it shows the poor performance of these methods, on the other hand, it implies that based on the estimated shape and scale parameters particularly by GM and EM, Weibull does not perfectly fit the historical wind speed in May-July. Similarly, with respect to chi-square test, the same conclusion can be drawn for GM, MM, and EPFM methods in month May, for GM and EM in month June, and for GM, MM, and EM in month July.

Table 3.11: Estimates of parameters and performance criteria based on different methods for monthly wind speed data at Khaf, $\boldsymbol{X^2}|_{\alpha=0.05}^{df=9} = \mathbf{16.919}$

Period	Methods	Parameters		R^2	RMSE	K-S	X^2
		c	k				
January	GM	6.3212	1.5550	0.967	0.0061	0.0091	9.83
	MM	6.0845	1.4682	0.979	0.0062	0.0143	8.46
	EM	6.0831	1.4656	0.979	0.0062	0.0099	10.31
	MLE	6.1212	1.4932	0.979	0.0062	0.0011	6.09
	EPFM	6.0662	1.4362	0.977	0.0063	0.0032	6.43
	GSO	5.9951	1.4356	0.981	0.0063	0.0020	5.17
February	GM	7.1106	1.2353	0.912	0.0166	0.0138	12.93
	MM	7.0737	1.1964	0.910	0.0164	0.0065	13.00
	EM	7.0721	1.1954	0.910	0.0164	0.0121	15.007
	MLE	7.1002	1.2009	0.917	0.0164	0.0039	12.82
	EPFM	7.1507	1.2508	0.908	0.0167	0.0060	14.27
	GSO	6.9531	1.0721	0.914	0.0178	0.0013	11.87
March	GM	6.7745	1.2675	0.895	0.0201	0.0320	16.21
	MM	6.6810	1.3972	0.896	0.0200	0.0390	16.10
	EM	6.6794	1.3951	0.897	0.0198	0.0393	15.90
	MLE	6.6544	1.3501	0.910	0.0153	0.0308	15.62
	EPFM	6.7247	1.4610	0.897	0.0188	0.0390	15.99
	GSO	6.7951	1.1674	0.912	0.0160	0.0295	15.30
April	GM	12.0541	2.1051	0.895	0.0252	0.0141	15.80
	MM	12.4723	2.0123	0.909	0.0144	0.0141	12.09
	EM	12.4732	2.0374	0.908	0.0143	0.0163	15.90
	MLE	12.4357	1.9820	0.912	0.0144	0.0139	11.58
	EPFM	12.4732	1.9205	0.906	0.0147	0.0128	12.75
	GSO	12.8013	1.9007	0.932	0.0105	0.0173	10.63
May	GM	13.4712	2.2574	0.828	0.0259	0.0264	18.05
	MM	14.0136	2.7324	0.856	0.0223	0.0132	17.82
	EM	14.0159	2.7215	0.853	0.0223	0.0219	16.66
	MLE	13.9539	2.7276	0.889	0.0222	0.0135	15.27
	EPFM	14.0127	2.7391	0.856	0.0225	0.0140	19.01
	GSO	14.0342	2.8851	0.900	0.0156	0.0138	13.97
June	GM	16.7064	2.8021	0.864	0.0186	0.0161	17.63
	MM	17.0465	3.9423	0.872	0.0171	0.0124	16.18
	EM	17.0525	3.9176	0.860	0.0190	0.0153	17.39
	MLE	16.9567	4.2756	0.889	0.0161	0. 0099	15.95
	EPFM	17.1661	3.4703	0.869	0.0209	0.0133	16.80
	GSO	16.00	4.2930	0.891	0.0196	0.0105	16.10

Table 3.12: Estimates of parameters and performance criteria based on different methods for monthly wind speed data at Khaf, $\boldsymbol{X^2}|_{\alpha=0.05}^{df=9} = \mathbf{16.919}$

Period	Methods	Parameters		R^2	RMSE	K-S	X^2
		c	k				
July	GM	14.1573	2.9830	0.837	0.0230	0.0183	20.03
	MM	14.2856	3.3317	0.865	0.0241	0.0194	19.80
	EM	14.2896	3.3133	0.846	0.0227	0.0214	19.65
	MLE	14.2773	3.3477	0.872	0.0165	0.0101	15.80
	EPFM	14.3376	3.0919	0.851	0.0200	0.0141	18.74
	GSO	14.3452	3.2742	0.880	0.0173	0.0108	15.06
August	GM	15.1021	2.901	0.886	0.0260	0.0105	15.84
	MM	15.4992	3.7394	0.916	0.0098	0.0112	13.95
	EM	15.5044	3.7168	0.915	0.0098	0.0134	11.28
	MLE	15.4536	3.8425	0.921	0.0098	0.0118	12.99
	EPFM	15.5923	3.3422	0.884	0.0200	0.0135	11.54
	GSO	15.6209	4.0029	0.937	0.0094	0.0126	9.29
September	GM	9.5761	1.7822	0.903	0.0110	0.0123	16.73
	MM	10.3599	1.9119	0.927	0.0097	0.0069	12.10
	EM	10.3589	1.9063	0.928	0.0093	0.0034	15.60
	MLE	10.2923	1.8157	0.935	0.0085	0.0135	13.78
	EPFM	10.3708	1.9963	0.918	0.0104	0.0137	11.34
	GSO	10.7350	1.8324	0.948	0.0072	0.0132	10.39
October	GM	9.1680	1.4168	0.896	0.0196	0.0102	15.11
	MM	9.9926	1.5375	0.901	0.0173	0.0135	15.96
	EM	9.9904	1.5344	0.885	0.0182	0.0116	15.54
	MLE	9.8303	1.4017	0.921	0.0090	0.0086	11.07
	EPFM	10.0404	1.6134	0.906	0.0152	0.0131	16.02
	GSO	10.3785	1.3879	0.934	0.0081	0.0090	12.21
November	GM	6.2532	1.3359	0.930	0.0098	0.0038	12.91
	MM	6.1460	1.2481	0.947	0.0091	0.0012	11.92
	EM	6.1446	1.2469	0.947	0.0091	0.0117	12.05
	MLE	6.1927	1.2684	0.945	0.0091	0.0136	11.18
	EPFM	6.1799	1.2796	0.934	0.0092	0.01070	15.52
	GSO	6.0514	1.1417	0.942	0.0099	0.0054	13.77
December	GM	6.2622	1.3105	0.906	0.0160	0.0121	13.24
	MM	6.3015	1.2620	0.914	0.0153	0.0112	11.01
	EM	6.300	1.2606	0.914	0.0153	0.0142	13.81
	MLE	6.3239	1.2648	0.913	0.0153	0.0127	10.05
	EPFM	6.3396	1.2984	0.906	0.0157	0.0097	12.29
	GSO	6.2145	1.1403	0.916	0.0159	0.0104	10.35

Though GSO and MLM perform more accurately and present the lower value of K-S test and chi-square compare to other methods in May-July, their larger value of RMSE and the smaller value of R-squared in these months compared to other months indicate that observed wind speed in May-July, notably in June and July does not perfectly follow the Weibull. Noted that K-S test's critical value of each month are presented in

Table 3.7, while the critical value of chi-square is obtained from the table in Appendix A. Figure 3.21 demonstrates how Weibull distribution fits the observed wind speed in September based on obtained parameters by numerical methods.

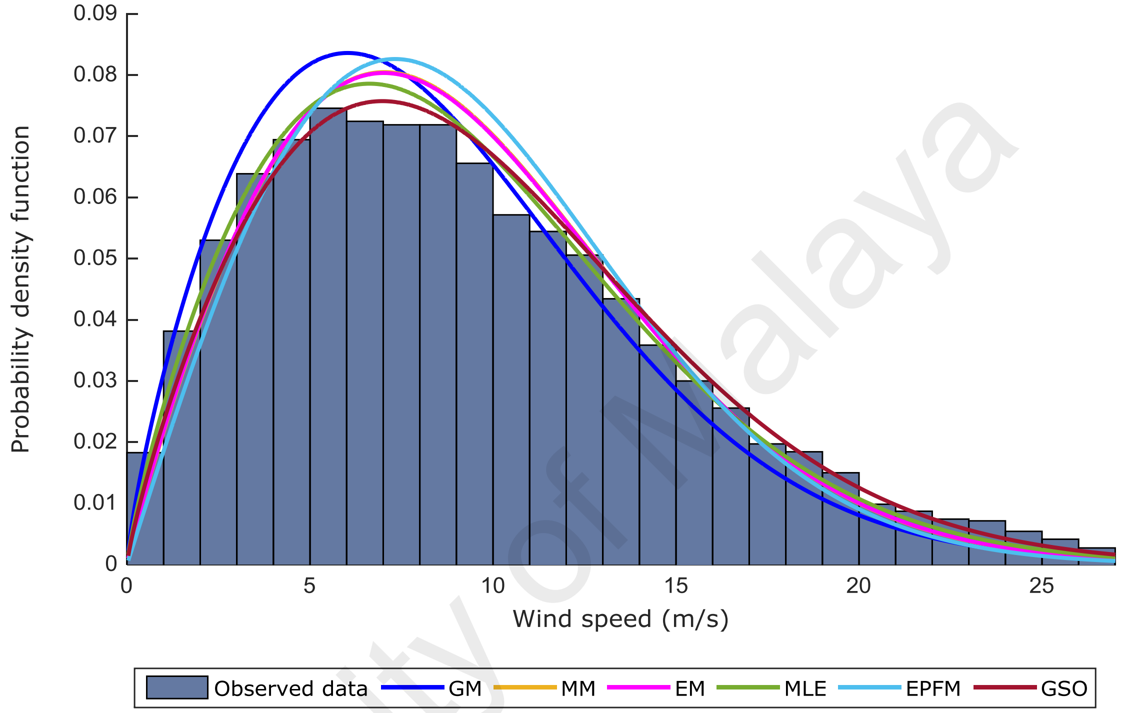
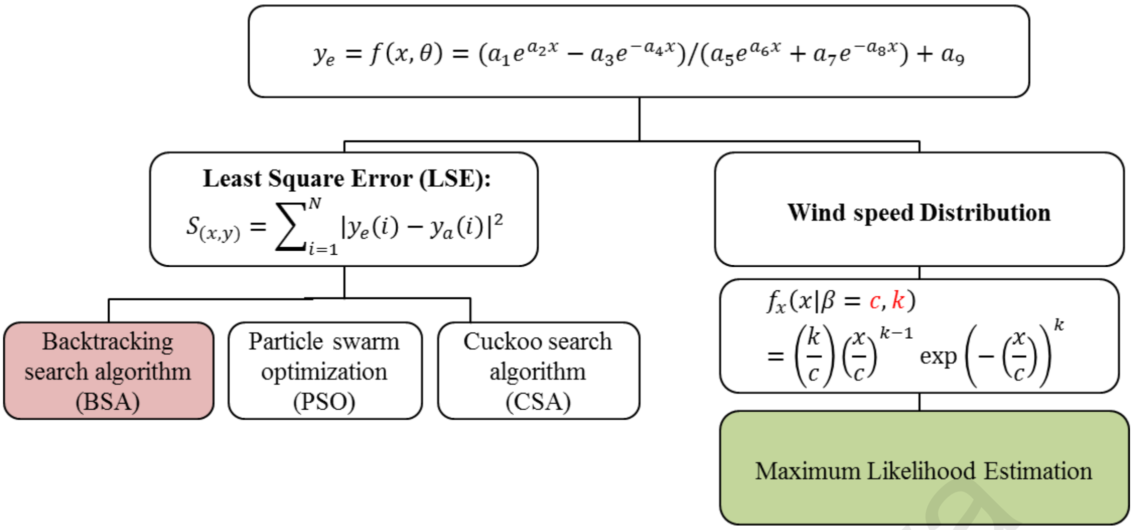


Figure 3.21: Observed frequency and probability density functions in September using different parameter estimation methods

3.7.3 Parameter Estimation of WTPC Modeling Based on LSE and MLE for Actual Data

To compare the performance of LSE and MLE in the parameters estimation of the MHTan, the objective functions Eqs. (3.2) and (3.59) are minimized by BSA. The 2-D scatter illustration shown in Fig. 3.2 depicts the performance of LSE and MLE in the spring season with 5 minutes interval. From MLE, the estimated vector based on data in spring is $\theta = (-7.89, -0.22, 88.34, -7.34, 0.05, -30.45)$. One can see that it is relatively different from the LSE estimation, which is $\theta = (-10.76, -0.26, 95.82,$ -7.96 -0.06,-33.73). The value of MAPE and RMSE obtained using MLE (10.5%, 169.4) are bigger as compared to LSE (1.98%, 31.4). Therefore, for further analysis, an average of 5 minutes data recorded in fall is analyzed and the similar pattern is observed. It shows that LSE has a greater performance than MLE, hence LSE will be used for parametric models estimation. In this study, the proposed MHTan is compared with polynomial order 6, polynomial order 7, four-parameter logistic function (4-PL), and five-parameter logistic function (5-PL). The vector parameters θ of the mentioned parametric models are depicted in Table 3.13, using LSE based on BSA algorithm in a 5-min interval data recorded in winter.



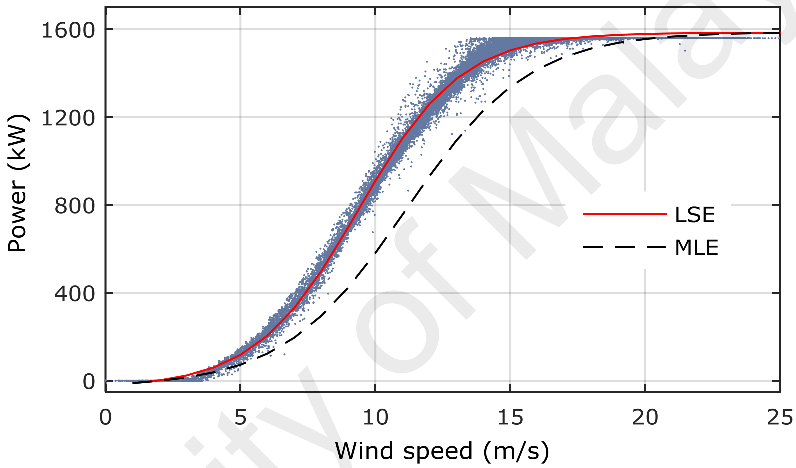


Figure 3.22: Comparison of MLE and LSE based on MHTan at 5-min averaged data in fall

Table 3.13: Estimated parameters value of parametric models using BSA based on 5-min data in winter

Model	Vector parameter θ								
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
Order6*	-12e-4	0.08	-2.17	24.05	-98.81	164.13	-80.33	–	–
Order7*	-6.9e-7	-6.0e-4	0.07	-1.89	21.81	-89.72	148.35	-72.29	–
4-PL*	-21.95	-0.30	-1.48	41e-4	–	–	–	–	–
5-PL*	1560	8.82	53.54	4.09	1560	–	–	–	–
MHTan	-1.0	-28.07	10.0	28.34	-.047	-28.57	-9e-4	28.09	52.18

*The formulas are presented in Appendix C.

3.7.4 Parameters Estimation of WTPC Modeling Based on Actual and Generated Data

This section used theoretical power curve data from the manufacturer to analyze the performance of parametric and nonparametric WTPC models. Firstly, three turbines having the different power-curve shape will be measured. Two turbines (Aeolos-50 (T_1) and Vestas V80 (T_3)) manufactured by Aeolos and Vestas, are horizontal-axis type, and have a power profile of 50 kW and 2MW, respectively, with hub heights of 24 m (Aeolos) and 80 m (Vestas). The third one is a UGE-4k vertical turbine (T_2) manufactured by Urban Green Energy Company with a power profile of 4kW and hub height of 4.6 m. For simulation of wind speed and power data, 720 data points for each data set are generated using normal distribution statistical method with mean (μ) equal to power data (obtained from manufacturers). The constant deviation for each types (Aeolos, UGE-4k and V80) are$\sigma_{\varepsilon} = 3$, $\sigma_{\varepsilon} = 0.25$, and $\sigma_{\varepsilon} = 120$, respectively.

The models used in this research are 4-PL and 5-PL, sixth- and seventh- degree polynomial. Additionally, two data mining algorithms, namely, bagging tree, and MLP (Kusiak, Zheng, & Song, 2009a) are employed. Grid partition (Gd) and subtractive clustering (Al-Shammari et al., 2015) are the two partition techniques to set up the ANFIS. To obtain the unknown parameters of the parametric WTPC modeling method, three optimization methods with a population size of 50 and 30 independent runs are executed in MATLAB. For the nonparametric models, a training set of 70%, together with the test and validation set of 30% are randomized and partitioned for both simulated and real data validation. Table 3.14 shows the details of all algorithms parameter. WEKA is used to develop MLP and bagging algorithms.

Table 3.14: Setting the parameters

Algorithm	Free parameter
BSA	Dim rate=0.9
CSA	Probability of an alien egg (Pa)= [0, 1] Distribution factor (ß)= 1.5
PSO	w_{max} = 0.9, w_{min} = 0.4, c_1=c_2=2
MLP	Set up as default
Bagging	Set up as default
Gd	numMfs=8, inmftype='trimf', outmftype='constant', epoch_n=50, Optim. method='hybrid'
Subtractive	Radii=0.5, epoch_n=50

Table 3.15: MAPE (%) of parametric & nonparametric models for actual data (winter and summer)

Model	Algorithm	5min		Hourly		3-hour	
		Summer	Winter	Summer	Winter	Summer	Winter
Order6	BSA	2.44	6.31	2.05	5.62	2.22	5.59
	CSA	2.50	6.30	2.05	5.62	2.22	5.60
	PSO	2.44	6.31	2.10	5.84	2.25	5.59
Order7	BSA	2.37	6.19	2.06	4.77	2.22	4.99
	CSA	2.61	6.20	2.04	4.91	2.21	4.97
	PSO	2.84	5.56	2.17	4.68	2.21	5.01
4-PL	BSA	1.97	5.28	1.88	5.19	2.23	5.55
	CSA	2.12	5.28	1.88	4.92	2.22	5.54
	PSO	1.97	5.35	1.87	4.99	2.23	5.56
5-PL	BSA	1.52	4.78	1.85	4.47	2.02	5.14
	CSA	1.52	4.78	1.85	4.47	2.04	5.13
	PSO	1.52	4.78	1.85	4.47	2.03	5.13
MHTan	BSA	1.57	3.83	1.62	3.93	2.01	4.90
	CSA	1.72	4.43	1.65	4.34	2.02	4.92
	PSO	1.60	4.60	1.60	3.92	1.99	4.95
Data mining	Bagging	2.18	4.96	2.11	5.32	2.19	5.16
	MLP	1.94	5.05	1.56	4.90	2.18	5.31
ANFIS	GD	3.18	8.32	2.65	7.23	2.68	6.72
	subtractive	1.90	4.55	1.53	4.61	2.12	4.99

Table 3.16: RMSE of parametric & nonparametric models for actual data (winter and summer)

Model	Algorithm	5min		Hourly		3-hour	
		Summer	Winter	Summer	Winter	Summer	Winter
Order6	BSA	38.11	38.77	33.63	33.67	38.32	34.19
	CSA	38.27	38.77	33.64	33.67	38.33	34.20
	PSO	38.11	38.77	34.23	36.04	38.36	34.21
Order7	BSA	36.77	38.73	33.93	32.77	37.31	32.29
	CSA	40.52	38.73	34.10	32.41	37.31	32.32
	PSO	43.62	41.02	35.04	32.73	37.50	32.30
4-PL	BSA	31.90	34.04	31.57	31.81	37.38	37.02
	CSA	33.50	33.83	31.55	31.87	37.38	37.02
	PSO	31.90	33.67	31.55	31.68	37.41	37.02
5-PL	BSA	28.97	30.16	30.01	28.96	35.73	34.23
	CSA	28.97	30.16	30.01	28.96	35.73	34.24
	PSO	28.98	30.16	30.01	28.96	35.74	34.23
MHTan	BSA	29.12	23.26	29.26	24.21	35.05	31.18
	CSA	29.84	28.03	29.76	27.26	35.60	31.86
	PSO	29.15	28.99	29.35	24.52	35.09	31.87
Data mining	Bagging	34.44	31.05	33.81	39.40	37.32	33.10
	MLP	32.02	33.33	28.90	31.93	37.39	33.56
ANFIS	GD	57.52	51.43	47.60	45.67	38.30	40.06
	subtractive	31.93	29.47	28.42	31.92	35.89	34.77

Table 3.17: Comparison of parametric & nonparametric models for T_1, T_2, and T_3 based on generated data

Model	Algorithm	Simulated data-MAPE (%)			Simulated data-RMSE		
		T_1	T_2	T_3	T_1	T_2	T_3
Order6	BSA	8.41	6.49	6.53	4.11	0.26	117.23
	CSA	8.44	6.49	6.53	4.13	0.26	117.23
	PSO	8.92	7.12	6.54	4.24	031	117.24
Order7	BSA	7.42	6.46	6.50	3.68	0.26	116.32
	CSA	7.43	6.70	6.51	3.67	0.26	116.45
	PSO	7.90	7.30	6.50	3.88	0.28	116.91
4-PL	BSA	8.70	5.74	6.36	4.11	0.24	114.39
	CSA	8.70	5.80	6.37	4.11	0.24	114.85
	PSO	8.65	5.75	6.36	4.14	0.24	114.38
5-PL	BSA	8.34	5.37	6.11	3.95	0.22	109.90
	CSA	8.34	5.37	6.11	3.95	0.22	109.90
	PSO	8.34	5.37	6.11	3.95	0.22	109.90
MHTan	BSA	6.85	4.96	5.93	3.45	0.21	107.85
	CSA	7.03	5.40	6.07	3.65	0.22	110.05
	PSO	8.70	5.06	5.99	4.17	0.21	108.90
Data mining	Bagging	6.79	4.88	6.12	3.35	0.20	117.14
	MLP	6.83	5.23	6.28	3.42	0.25	114.56
ANFIS	Gd	6.95	5.82	6.50	3.68	0.26	119.50
	Subtractive	6.82	5.64	6.01	3.44	0.24	108.98

The performance of the parametric and nonparametric models in terms of RMSE and MAPE for the real data sets representing 5-min, hourly, and 3-hour average values of the data measured in the winter and summer as well as for the simulated data is given in Tables 3.15 and 3.16. Similarly, the simulated data is shown in Table 3.17. Note that there is a difference in error values for data measured in winter and summer. This is probably caused by temperature different during the wind turbine operation. From the achieved data, it can be concluded that the wind speed is flowing the Weibull distribution in winter while normal distribution in summer. The performance of all methods along with the overall performance is ranked for both simulated and actual data based on statistical errors, as presented in Tables 3.18 to 3.19. From the tables, it can be clearly seen that the proposed MHTan model utilizing the BSA-based algorithm has the best overall performance. However, its performance is yet promising even by applying PSO and CSA to estimate vector control. The reason MHTan achieves satisfactory results is attributed to a combination of four exponential functions with different exponents in the MHTan structure. In other words, unequal exponential functions in MHTan structure make this model highly flexible for fitting the power curve. Although, the value of exponents is very close together as will be seen later, this slight difference considerably influences the shape of MHTan. Figure 3.23 depicts the changes in shape of MHTan versus four exponents a_2, a_4, a_6, and a_8.

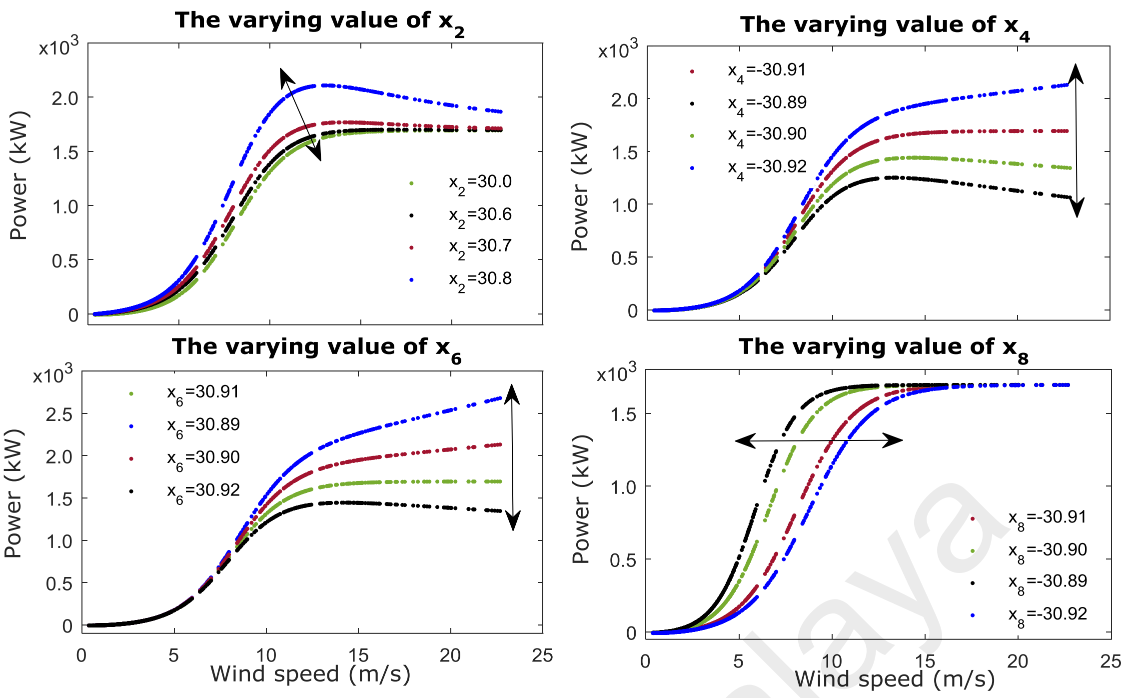


Figure 3.23: Influence of parameters $\boldsymbol{a_2}, \boldsymbol{a_4}, \boldsymbol{a_6}$, and $\boldsymbol{a_8}$ on MHTan behavior

Table 3.18: Performance ranking of parametric and nonparametric models for actual and simulated dataset (MAPE)

Model	Algorithm	5min		Hourly		3-hour		Simulated data			Average rank	Overall performance
		Summer	Winter	Summer	Winter	Summer	Winter	T1	T2	T3		
Order 5	BSA	11	14	10	14	10	14	11	13	12	12.11	13
	CSA	12	13	10	14	10	15	12	13	12	12.33	14
	PSO	11	14	12	15	12	14	15	15	13	13.44	16
Order 7	BSA	10	11	11	7	10	5	7	12	10	9.22	9
	CSA	13	12	9	9	9	4	8	14	11	9.88	11
	PSO	14	10	14	6	9	6	9	16	10	10.44	12
4-PL	BSA	7	8	8	12	11	12	14	8	8	9.77	10
	CSA	8	8	8	10	10	11	14	10	9	9.77	10
	PSO	7	9	7	11	11	13	13	9	8	9.77	10
5-PL	BSA	1	5	6	4	3	8	10	5	5	5.22	5
	CSA	1	5	6	4	5	7	10	5	5	5.33	6
	PSO	1	5	6	4	4	7	10	5	5	5.22	5
MHTan	BSA	2	1	4	2	2	1	4	2	1	2.11	1
	CSA	4	2	5	3	3	2	6	6	4	3.88	3
	PSO	3	4	3	1	1	3	14	3	2	3.77	2
Data mining	Bagging	9	6	13	13	8	9	1	1	6	7.33	8
	MLP	6	7	2	8	7	10	3	4	7	6.00	7
ANFIS	GD	15	15	15	16	13	16	5	11	10	12.88	15
	subtractive	5	3	1	5	6	5	2	7	3	4.11	4

Table 3.19: Ranke table based on RMSE for both actual and generated data

Model	Algorithm	5min		Hourly		3-hour		Simulated data			Average rank	Overall performance
		Summer	Winter	Summer	Winter	Summer	Winter	T_1	T_2	T_3		
Order 5	BSA	12	12	9	13	14	9	10	6	14	11	15
	CSA	13	12	10	13	15	10	11	6	14	11.55	16
	PSO	12	12	14	14	16	11	14	8	15	12.88	17
Order 7	BSA	11	11	12	12	7	4	7	6	10	8.88	12
	CSA	14	11	13	10	7	6	6	6	11	9.33	13
	PSO	15	13	15	11	12	5	8	7	12	10.88	14
4-PL	BSA	6	10	8	6	9	15	10	4	7	8.33	10
	CSA	9	9	7	7	9	15	10	4	9	8.77	11
	PSO	6	8	7	5	11	15	12	4	6	8.22	9
5-PL	BSA	1	5	6	4	4	12	9	3	4	5.33	4
	CSA	1	5	6	4	4	13	9	3	4	5.44	5
	PSO	2	5	6	4	5	12	9	3	4	5.55	6
MHTan	BSA	3	1	3	1	1	1	4	2	1	1.89	1
	CSA	5	2	5	3	3	2	5	3	5	3.66	2
	PSO	4	3	4	2	2	3	13	2	2	3.88	3
Data mining	Bagging	10	6	11	15	8	7	1	1	13	8.00	8
	MLP	8	7	2	9	10	8	2	5	8	6.55	7
ANFIS	GD	16	14	16	16	13	16	7	6	16	13.33	18
	subtractive	7	4	1	8	6	14	3	4	3	5.55	6

From the results, it can be seen that parametric models obtained better results than the nonparametric models, in terms of accuracy, as shown in Tables 3.18 and 3.19. However, one drawback is that not many tools exist to filter the outlier's original data, which is crucial in this method. It is worth mentioning that, in artificial neural network viewpoint, parameter selection has been always a rigorous task because there is no decisive rule to obtain the optimum parameters (Ramirez-Rosado, Fernandez-Jimenez, Monteiro, Sousa, & Bessa, 2009). Several experiments have been performed to determine which structures produce better results. The value of 'radii' and epoch number will heavily affect the ANFIS subtractive clustering application. The influence range of input and output cluster is indicated by the value of 'radii' while the training iteration is indicated by the epoch number. The overall performances of ANFIS utilizing various values of 'radii' and RMSE's 5 minutes epoch number data in winter are shown in Table 3.20. It can be concluded that no obvious improvement can be seen at higher epoch number. As a result, the number is limited to 50. During the analysis, no statistical metrics like MAPE is used. In addition, different data sets including simulated and actual data lead to the same performance for the presented structures in Table 3.20.

Table 3.20: RMSE results of ANFIS using subtractive clustering based on 5-min data in winter

Epoch Number	Radii [influence of a cluster in input - influence of a cluster in output]						
	[0.8-0.2]	[0.7-0.3]	[0.5-0.4]	[0.4-0.6]	[0.4-0.1]	[0.8-0.1]	[0.5-0.5]
10	53.10	51.92	36.13	47.84	39.72	48.68	33.49
20	50.80	49.41	34.98	44.13	38.42	46.93	30.82
30	49.63	47.88	33.67	43.62	36.30	46.01	30.43
40	48.10	47.06	32.04	41.19	34.96	44.76	30.10
50	47.95	45.35	31.71	40.18	36.11	44.01	29.47

Different structures were used during observation of ANFIS performance utilizing a grid partition. It was evaluated by different epoch numbers, number of membership, as well as different membership functions such as Gaussian, trapezoidal, and triangularshaped. For the sake of conciseness, only performance evaluation of ANFIS using subtractive clustering is illustrated in Table 3.20. The reason why this is choosen is due to the superior results of ANFIS based on clustering subtraction when compared to grid partition. Eventually, ANFIS is developed with selected parameters in Table 3.14.

For performance observation, ANFIS based subtractive clustering has shown better results compared to other nonparametric model. This is due to the capability of subtractive clustering as a one-pass algorithm in estimating the number of clusters and extracting the fuzzy rules. As compared with other models, data mining by bagging algorithm maps the relationship between wind speed and the power output of turbines T_1 and T_2 well. All in all, for both T_1 and T_2, the main reason for the better performance of the nonparametric models over the parametric models is that turbine power curve usually drops when the wind is over the rated value. However, nonparametric models still suffer the black-box problem. Alternatively, combining both nonparametric and parametric models might produce adequate results and enhance the facilitation of wind turbines monitoring.

The parameters of MHTan obtained by BSA, CSA, and PSO for winter and summer at 5-min, hourly, and 3-hour interval are listed in Table 3.21. The performance of MHTan model on four turbines with different power curve shapes as compared to power curves provided by manufacturers are shown in Figures 3.24 to 3.26. Meanwhile, the mismatching between the observed data and the data taken from the turbine WD77 manufacturer as illustrated in Fig.35 has become the motivation for improved wind turbine power curve modeling.

Table 3.21: Estimated parameters value of MHTan based on real data insummer

	Period	Algorithm	Vector parameter θ								
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
5-min	Summer	BSA	-11.39	-28.63	1.79	28.33	-0.49	-28.97	-12e-4	28.34	-62.49
		CSA	-18.90	-27.86	13.56	27.60	-2.09	-28.21	-82e-4	27.60	-49.12
		PSO	-11.87	-25.12	3.81	25.0	-1.23	-25.62	-49e-4	25.02	-39.35
	Winter	BSA	-1.0	-28.07	10.0	28.34	-.047	-28.57	-9e-4	28.09	52.18
		CSA	-19.0	-26.19	13.89	25.87	-1.98	-26.51	-76e-4	25.86	-37.6
		PSO	-13.0	-30.14	7.86	29.62	-0.84	-30.24	-40e-4	29.61	-44.30
1-Hour	Summer	BSA	-9.26	-31.45	2.98	31.28	-0.56	-31.86	-28e-4	31.29	-55.0
		CSA	0.50	30.03	-0.28	-30.15	-.05	29.53	3e-4	-30.12	-24.98
		PSO	-2.93	-27.77	1.04	27.53	-7e-4	-27.54	-0.16	28.12	-55.0
	Winter	BSA	-3.25	-32.21	16.65	34.99	-0.29	-32.82	-17e-4	32.20	-31.0
		CSA	-19.99	-28.14	11.80	28.02	-1.81	-28.6	-0.01	28.03	-48.86
		PSO	-12.0	-30.76	6.72	30.58	-1.14	-31.21	-49e-4	30.58	-39.25
3-Hour	Summer	BSA	-2.57	-28.27	0.68	28.07	-0.10	-28.63	-6e-4	28.08	-54.90
		CSA	0.32	30.71	-0.37	-30.61	-0.04	30.15	3e-4	-30.70	-30.0
		PSO	-19.0	-30.78	6.33	30.58	-0.83	-31.13	-50e-4	30.59	-61.43
	Winter	BSA	-10.14	-30.89	2.05	28.55	-0.10	-29.10	-10e-4	28.54	-51.83
		CSA	-1.87	-32.0	10.0	34.32	-9e-4	-31.98	-0.09	32.54	-52.48
		PSO	2.50	28.50	-0.30	-31.50	0.01	30.96	1e-4	-31.51	-50.05

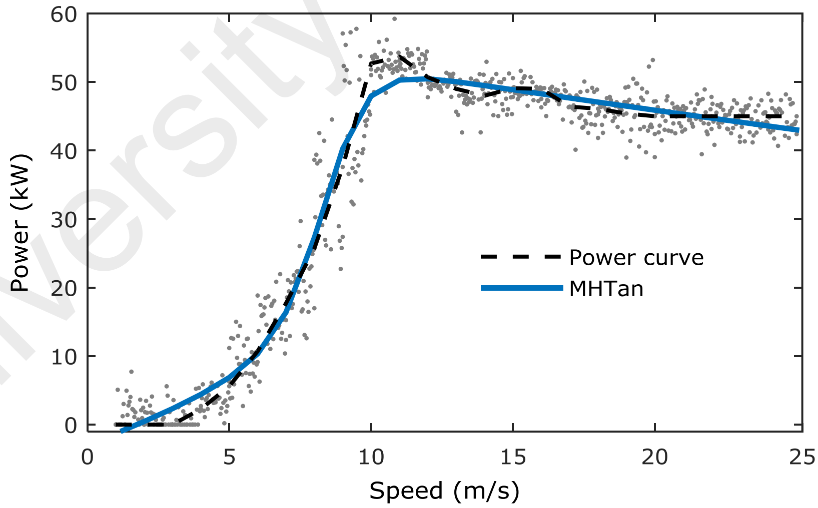


Figure 3.24: Approximated power curve by MHTan for the power data generated from turbine Aeolos-50 (stall-based turbine).

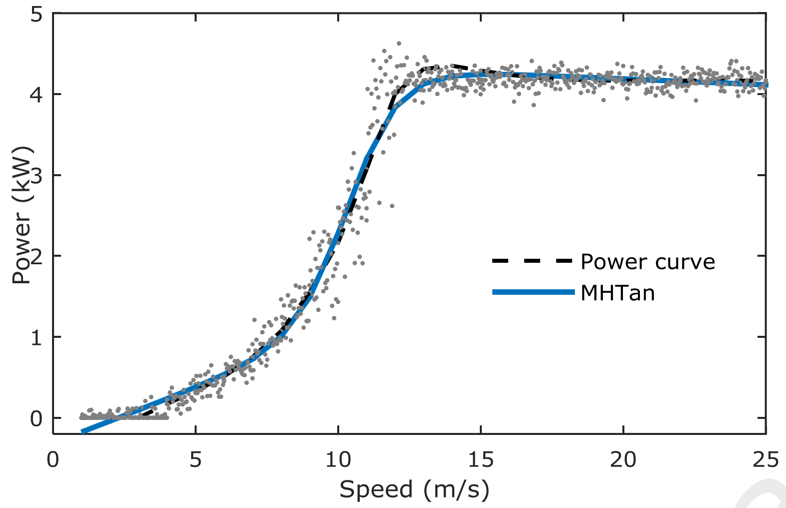


Figure 3.25: Approximated power curve by MHTan for the power data generated from turbine UGE-4 (vertical axis turbine)

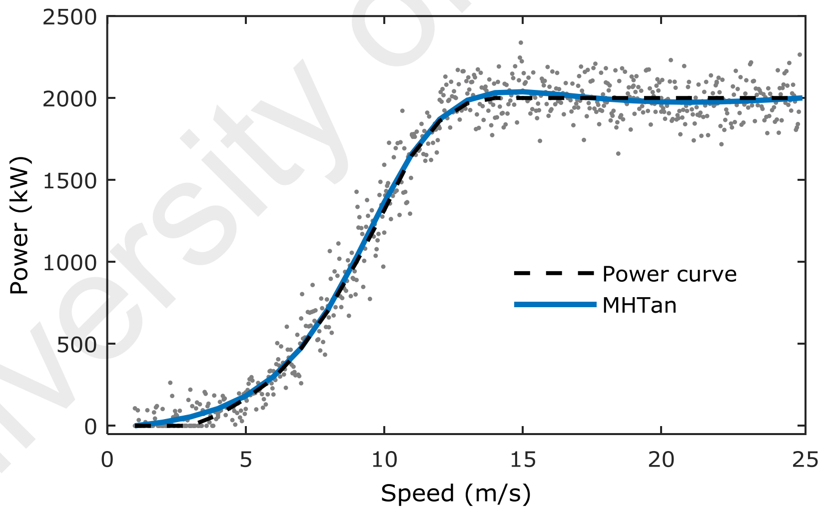


Figure 3.26: Approximated power curve by MHTAn for the power data generated from turbine V80 (pitch-based turbine)

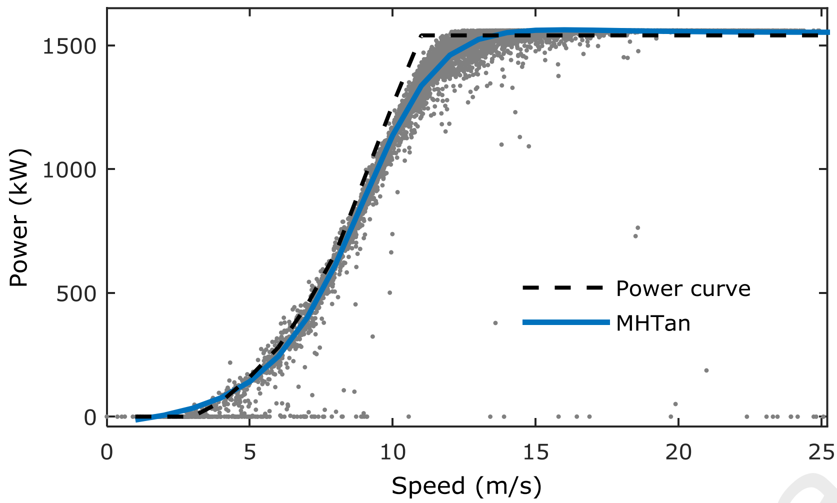


Figure 3.27: The scatter plot of the theoretical power curve and approximated power curve by MHTan based on observed wind speed and power (empirical power curve)

3.7.5 Sensitivity Analysis of Optimization Algorithms

The performance of parametric methods is affected by the control parameters of the optimization algorithms. To produce the better results, it should be finely tuned. For sensitivity analysis, three BSA parameters can be considered namely population size, mixrate and the maximum iteration. Mixrate values of 0.1-1.0 (step size of 0.10), maximum iterations of 1000 – 2000 (step size of 250) and population size of 10 – 50 (step size of 10) are considered for the purpose of sensitivity analysis. For analysis of mixrate effect on the optimal result, population size is set to 50, while maximum iteration number is set to 2000. Then, the optimization procedure is run by varying the parameter of mixrate. Table 3.22 shows the RMSE of MHTan for different values of mixrate. It illustrates that the BSA reaches the better optimal value by selecting relatively large value for the mixrate as it is equal to 0.9 in this case.

Table 3.22: RMSE statistics of MHTAn using BSA with different mixrate values based on 5-min data in winter (population size=50 and maximum iteration=2000)

mixrate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Minimum	22.80	23.08	22.07	21.95	21.92	21.89	21.83	21.81	21.44	21.82
Average	26.78	26.43	25.72	26.07	25.41	25.63	26.24	25.30	23.26	24.98
Maximum	31.15	30.54	30.81	31.13	30.56	29.97	29.89	29.68	27.05	28.90
Standard deviation	2.26	1.82	2.32	1.80	2.42	2.57	2.49	2.50	2.30	2.39

The analysis of the effect of population size on optimal results is conducted by setting the mixrate and the maximum iteration to 0.9 and 2000 respectively, and the population size from 10 to 50. The optimization is run 30 times and the statistical indices of the results are shown in Table 3.23. Based on the average values of optimal results, it is concluded that the increasing population size leads to a better result.

Table 3.23: RMSE statistics of MHTAn using BSA with different population size values based on 5-min data in winter (mixrate=0.90 and maximum iteration=2000)

Population size	10	20	30	40	50
Minimum	24.43	23.08	23.56	22.18	21.44
Average	27.36	25.99	26.08	24.81	23.26
Maximum	29.97	28.65	28.95	27.87	27.05
Standard deviation	2.71	2.51	2.64	2.81	2.92

Although it is clear that the increasing maximum iteration will help the optimizer produces better optimal, the simulation is done to show the degree of effectiveness of this parameter on the optimal results. Based on the results in Table 3.24, the average optimal value shows about 2 % decrease within 30 trials when the maximum iteration is changed from 1000 to 1250. These decreases are about 5 %, 8 %, and 15 % for increasing maximum iteration to 1500, 1750, and 2000, respectively. It’s worth mentioning that since the BSA performs better than the PSO and CSA, only the sensitivity analysis of BSA is presented in this study.

Table 3.24: RMSE statistics of MHTAn using BSA with different number of iteration based on 5-min data in winter (mixrate=0.90 and population size=50)

Max. iteration	1000	1250	1500	1750	2000
Minimum	24.67	23.46	22.82	22.20	21.44
Average	27.28	26.68	25.79	25.19	23.26
Maximum	29.98	29.37	28.05	27.49	27.05
Standard deviation	2.60	2.86	3.01	2.90	2.92

3.7.6 On-line Monitoring by Residual Approach and Control Charts

MHTan model characterizes wind turbine power in normal conditions, thus, it can serve as an on-line wind farm power generation profile. It can be employed to detect if a wind turbine is deviating from the expected performance and then allowing for troubleshooting on fault conditions. Indeed, monitoring of the wind turbine has practical importance to reduce maintenance costs and improve operational efficiency and reliability (Park, Lee, Oh, & Lee, 2014). The residual control chart technique (Montgomery, 2009) is applied to assess residual between predicted power value by MHTan model and observed power. The purpose of control chart approach is to monitor residuals and their variations for detecting abnormal behaviors of the wind turbine. The 5-min data collected in August including 7816 observations are used to build the control chart. Observations are randomly separated into training and testing data with 5862 and 1954 data points respectively. Residual ε of each individual observation, standard deviation (σ_{Train}) and mean (μ_{Train}) of ε for training data are computed and expressed as follows:

$$\varepsilon = y_a - y_e$$

$$\mu_{Train} = \frac{1}{n}\sum_{i=1}^{n_{Train}} (y_a(i) - y_e(i)) \tag{3.69}$$

$$\sigma_{Train} = \left[\frac{1}{n-1}\sum_{i=1}^{n_{Train}} (y_a(i) - y_e(i))^2\right]^{1/2}$$

where y_a is actual power, y_e is predicted power by MHTan and n is the number of point in training data set. Once the mean (μ_{Test}) and standard deviation (σ_{Test}) of test data set are similarly computed, control limits can be derived as follows (Marvuglia & Messineo, 2012):

$$UCL_1 = \mu_{Train} + \varrho \frac{\sigma_{Train}}{\sqrt{n}}$$
$$LCL_1 = \mu_{Train} - \varrho \frac{\sigma_{Train}}{\sqrt{n}} \quad (3.70)$$

where UCL_1 and LCL_1 indicate upper and lower control limits respectively, ϱ is an integer multiple of the control limits, which regularly fixed as 3 and n is the number of samples in test data set which can be adjusted. Noted that the sensitivity of the control chart to data variability depends on the value of n and ϱ. Decreasing value of ϱ, results in strict upper and lower limiter. While the higher value of ϱ reduces the sensitivity to variability in observations and consequently increases the number of anomalies. Apparently, the influence of n is inverse of the parameter ϱ. If μ_{Test} is above UCL_1 or belowLCL_1, the power generation process at the sampling time $y_{TestSet} = [y_a(i), y_e(i)]$ is considered abnormal, otherwise it is considered efficient. Similarly, the control chart for σ_{Test}^2 can also be calculated as Eq. (3.71) to detect anomalies (Montgomery, 2009):

$$UCL_2 = \frac{\sigma_{Train}^2}{n-1} + \chi_{\alpha/2,n-1}^2$$
$$LCL_2=0 \quad (3.71)$$

where $n-1$ is the degrees of freedom of the chi-square distribution. Parameter α can be set to adjust the sensitivity of the control chart to variability of the data. Here, the value of parameter α was set to 2. If σ_{Test}^2 is above UCL_2, the observed power at the sampling time $y_{TestSet} = [y_a(i), y_e(i)]$ is considered anomaly and will be removed. For the purpose of indicating the variation of residuals in data set is 0, UCL_2 is set to 0. In this case, the current power matches the reference power in the normal status. Figure 3.28 illustrates power curve derived from the test data set and also anomalies detected by MHTan when in Eq. (3.71) n is set to 2. In this case, 258 observations out of 1954 are detected by MHTan. While the number of detected anomalies decreased to 188 when n is set to 1.

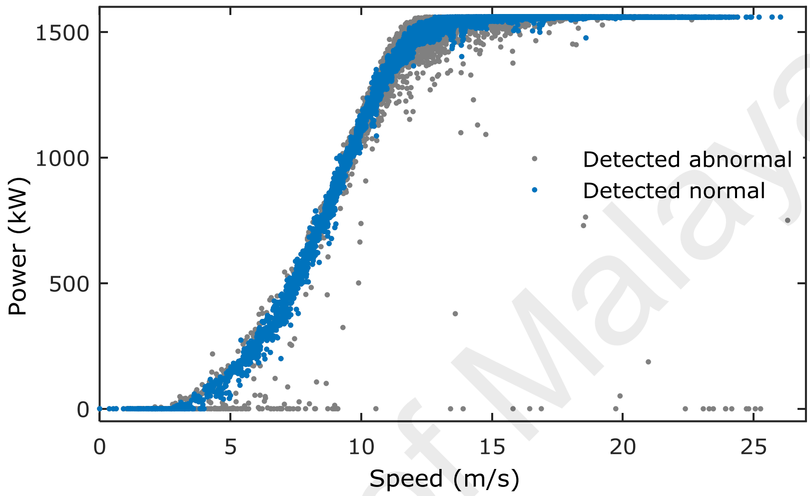


Figure 3.28: Anomalies detected by MHTan

3.8 Summary

An enhanced parametric model, called MHTan including nine unknown parameters, was presented in this chapter to model the empirical wind turbine power curve. Two approaches were employed to determine the unknown parameters of MHTan, one based on the optimization algorithms and the other one based on maximum likelihood method. Based on the simulation results, among three optimization algorithms, BSA was the best though the obtained results by the other optimizers were promising too. Despite the good performance of GSO in the estimation of wind speed parameters, model based on optimization algorithms performed better than the model based on MLE and wind speed distribution. To validate the performance of the MHTan, the actual data along with three simulated data sets representing yaw-, stall-, and pitch-controlled wind turbine are employed in this chapter and the results were compared with several parametric and nonparametric models. The analytical results not only affirmed the outperforming of MHTan but also its applicability in on-line monitoring of the wind turbine.

CHAPTER 4: DIRECT AND INDIRECT WIND POWER PREDICTION

4.1 Introduction

Prediction of the future wind power generated based on two methods, the direct and indirect prediction is investigated in this chapter. Forecasted wind speed is an important prerequisite to indirect prediction methods. To do so, three statistical methods are employed and the best one is chosen for prediction from few minutes to an hour. Then, the proposed MHTan in Chapter 3 is used to convert the forecasted wind speed to wind power. Moreover, six nonparametric methods are employed to present direct prediction. Similarly, the best method is determined to forecast wind power for longer horizons. In this chapter, in addition, a new feature selection (FS) technique is developed to reduce the number of initial input features without losing information. To validate the proposed FS technique, it is compared with several linear and nonlinear FS models. The programming code is written in MATLAB and executed on a personal computer with Intel Pentium 2.66 GHz processor and 4 GB RAM. Two out of six data-driven approaches which are M5rule and random forest are conducted by WEKA.

4.2 Wind Speed Forecasting

Wind speed prediction models are mainly categorized into two groups, time-series based and the weather based. In the latter physical data, i.e. topography information and temperature are applied to predict the wind speed. These models do not achieve reliable results in the short-term prediction thus this research mainly focuses on statistical approaches to forecast the wind speed. Time-series based models use recursive algorithms to forecast wind speed (Palomares-Salas, De La Rosa, Ramiro, Melgar, Aguera, et al., 2009; Sterba & Hilovska, 2010).

4.2.1 Double Exponential Smoothing (DES)

This technique does not assign the equal weights to the past observations but as the observations get older their weights exponentially decrease. In other words the basic concept of this method is that relatively more weight is given to the recent observation than the one given to older observations. (Taylor & McSharry, 2007). DES includes two constant parameters, η and λ, and attempts to calculate the estimated trend and level at time t which is expressed as follows:

$$
\begin{gathered}
S'_t = \eta y_t + (1 - \eta)(S'_{t-1} + b_{t-1}) \\
b_t = \lambda(S'_t - S'_{t-1}) + (1 - \lambda)b_{t-1} \\
\hat{y}_{t+1} = S'_t + b_t \\
S'_1 = y_1 \\
b_1 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3
\end{gathered}
\tag{4.1}
$$

where S'_t represents the smoothed value at time t, $\hat{y}_{t+1}$ is the predicted value, b_t is the best estimation of the trend for the time t, , η and λ are smoothing factors which vary from 0 to 1. It is noted that the unknown parameters of DES are obtained by BSA. The detail of BSA is provided in section 3.3.1.

4.2.2 Auto-Regressive Moving Average (ARMA)

In this technique, in order to estimate the future value, the past observations as well as the residual value from the past prediction are incorporated. This model, known as ARMA (*p, q*), combines the auto-regressive and moving average model by simply adding them together which is defined as follows:

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{4.2}$$

where p and q are the order of the auto-regression (AR), and the order of the moving average (MA) respectively, φ_i represents the AR coefficient, while θ_i is the MA coefficient, L is the lag operator and ε_t is the uncorrelated innovation process with zero mean and variance σ^2.

4.2.3 Persistence Method

Persistence algorithm is the simplest way of producing a prediction. Indeed, this method employs the value in prior step $(t-1)$ to forecast the value at the next step t. There is no required any parameter setting or exogenous variables in this model. Nevertheless, it yields better result than NWP in short-term prediction (Bludszuweit, Domínguez-Navarro, & Llombart, 2008).

4.3 Wind Power Prediction through Power Curve

One of the ways to predict the future output power is using WTPC which is called indirect power forecasting. In this technique, firstly, the wind speed is forecasted, then it is converted to power through wind power curve. Although many researchers have used theoretical power curve given by manufacturers, it produces some inaccuracies because it is an ideal power curve. To alleviate this problem, the developed parametric wind power curve modeling (MHTan) is applied to obtain the future power. MHTan is explained in detail in chapter 3 and its effectiveness is compared with other parametric and nonparametric models.

4.4 An Intelligent Feature Selection Technique

Feature selection technique is of vital importance in machine learning and data mining wherein a huge amount of data might be involved. Most of these data are usually noisy and carry the redundant information. The principle objective of the feature selection is to discover and remove the irrelevant and redundant input data and then select the best subset of attributes. The selected subset of inputs certainly represents the important information of the initial data. Feature selection techniques are broadly classified into three categories: filter methods, embedded methods and wrapper methods. Filter methods attempt to evaluate the redundancy of the initial data but not the interactions between data themselves. Embedded methods conduct the selection of features as a part of learning procedure. Indeed, variable selection and training process cannot be separated. Wrapper approaches, unlike the filter methods, evaluate a subset of variable which results in detecting the possible relationship between variables (Maldonado, Carrizosa, & Weber, 2015). A big majority of the feature selection methods have difficulty in choosing the redundant and irrelevant data. To address this issue, a nonlinear feature selection technique based on the mutual information (MI) is introduced in this study. Mutual information is closely tied with the entropy and can mitigate the problem of overestimation of the feature significance. The proposed feature selection technique includes two stages. In the first stage, irrelevant data are captured and removed based on mutual information and then in the next stage, the neural network is applied to filter out the redundant data. The following expounds the concept of information theory.

The entropy of random variable regardless of discrete or continuous presents the average amount of information that can be learned from the random variable X and this is a measure of its uncertainty. The entropy of discrete random variable $X = (X_1, X_2, \dots, X_n)$ denoted by (X), is defined as:

$$H(X) = -\sum_{i=1}^{n} P(X_i)\log\big(P(X_i)\big) = -\mathbb{E}\big[\log\big(P(X)\big)\big] \qquad (4.3)$$

where $P(X)$ is the probability function which is obtained by division of the number of samples with value of X_i to the total number of samples (n). Although the base of logarithm function is two which results in $H(X)$ varying between 0 and 1, choosing the base is arbitrary since it only changes the unit of entropy. If X is continuous random variable with probability function of $P(X)$ then $H(X)$ is expressed as follows:

$$H(X) = -\int P(X)\log\big(P(X)\big)\,dX \tag{4.4}$$

If X and Y are two continuous random variables, then the joint entropy of X and Y is defined as:

$$H(X,Y) = -\iint P(X,Y)\log\big(P(X,Y)\big)dXdY \tag{4.5}$$

where $P(X,Y)$ is the joint probability distribution of variables and $H(X,Y)$ presents the total amount of uncertainty of two random variables X and Y. Conditional entropy measures the remaining uncertainty of variable X when the value of Y is known. Conditional entropy is typically equal to or greater than zero, but it is equal to entropy of variable X and Y when both variables are absolutely independent. Conditional entropy denoted by (X,Y) is expressed as:

$$H(X|Y) = -\iint P(X,Y)\log\big(P(X|Y)\big)dXdY \tag{4.6}$$

The relationship between conditional entropy and joint entropy is known as chain-rule and defined as:

$$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X) \tag{4.7}$$

This states that the total uncertainty of variable X and Y is equal to the uncertainty of X plus the remaining entropy of Y when X is known.

Mutual information is a measure of the amount of information that one variable contains about another variable and it is expressed as follows:

$$\begin{aligned} I(X;Y) &= \iint P(X,Y)\log\left(\frac{P(X,Y)}{P(X)P(Y)}\right)dXdY = H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) = I(Y;X) \end{aligned} \tag{4.8}$$

Mutual information has two important properties. First, it is capable of measuring any kind of relationship between variables. Second, it is invariant to space transformation due to the fact that logarithm function used in Eq. (4.8) is non-dimensional. Venn diagram in Figure 4.1 illustrates mutual information of two variable X and Y.

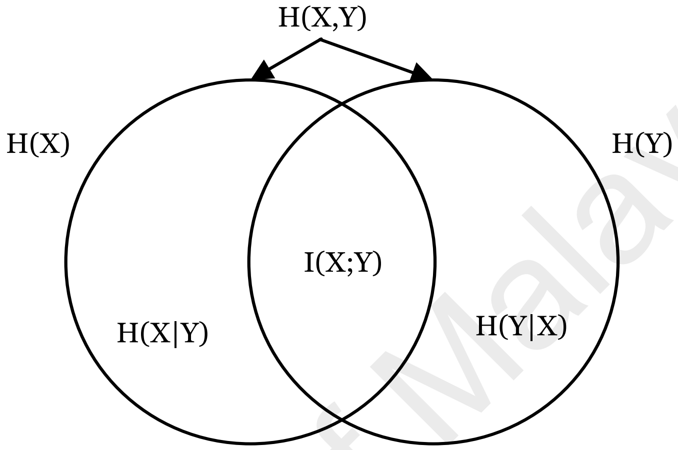


Figure 4.1: Graphical representation of the conditional entropy and the mutual information

In the process of wind power forecasting, wind power is largely a function of wind speed, temperature, wind direction, and humidity. It is well known that the power produced in time t depends not only on the metrological variables at time t but also on their past values and even the past values of the power generated which is expressed as follows:

$$\begin{aligned} WP(t) = f(&WS(t), WS(t-1), \dots, WS(t-NL_{WS}), T(t), T(t-1), \dots, \\ &T(t-NL_T), WD(t), WD(t-1), \dots, WD(t-NL_{WD}), H(t), \dots, \\ &H(t-1), \dots, H(t-NL_H), WP(t-1), \dots, WP(t-NL_{WP})) \end{aligned} \quad (4.9)$$

where $WS(t), T(t), WD(t),$ and $H(t)$ are the present wind speed, temperature, wind direction and humidity respectively, $WP(t)$ presents the forecasted power at time t, NL_{WS} denotes the lag length for the wind speed and similarly NL_T, NL_{WD}, NL_H, and

NL_{WP} presents the lag order of other variables. The above-mentioned variables, however, have a strong relationship with wind power but it is not feasible nor efficient to apply all of them into the prediction machine as inputs. Assuming that 25 lagged values of the variables ($NL_{WS} = NL_T = NL_{WD} = NL_H = NL_{WP}$ =25) are employed to predict the wind power, this constitutes 128 inputs inclusive of 100 past values of power, humidity, wind direction, temperature, and wind speed along with the present values of humidity, temperature, and wind speed as inputs to prediction machine. Such a massive amount of data will slow down the forecasting process. Moreover, it leads to a poor performance and in turn overfitting the training data when the engine forecast is a machine-learning algorithm. In fact, these variables, however, are greatly important for the prediction of wind power, but only those inputs representing considerable influence on the output power should be selected.

Let $\mathrm{X} = \{\mathrm{WS(t)}, \ldots, \mathrm{WS(t - NL_{WS})}, \mathrm{T(t)}, \ldots, \mathrm{T(t - NL_T)}, \mathrm{WD(t)}, \ldots, \mathrm{WD(t - NL_{WD})}, \mathrm{H(t)}, \ldots, \mathrm{H(t - NL_H)}, \mathrm{WP(t - 1)}, \ldots, \mathrm{WP(t - NL_{WP})}\}$ denote a vector including all input data and let $Y = WP(t)$ present a output feature. According to Eq. (4.8), the developed feature selection technique in the first step attempts to compute the mutual information between the each individual input and the target feature. According to Eq. (4.8), the developed feature selection technique in the first step attempts to compute the mutual information between each individual input $\forall X_i (1 < i < \mathrm{NL_{WS}} + \mathrm{NL_T} + \mathrm{NL_{WD}} + \mathrm{NL_H} + \mathrm{NL_{WP}})$ and the target feature. For example, $MI(X_2; Y)$ illustrates the mutual information between the present output power and one past value of wind speed. Input data are sorted in descending order of mutual information value in which the greater value of mutual information presents the stronger relationship between each individual input X and Y.

The variables having a higher value of mutual information than the given *TH* exhibiting the significant influence on the target feature remains for the second stage and form the subset $XS \subset X$. Input features with the mutual information lower the chosen *TH* are considered as irrelevant inputs and will be removed. Since the threshold TH is set by user, it must be noted that the lower *TH* value may include many irrelevant and redundant features resulting in huge computational cost, while lots of important information might be missed due to the high value of *TH*. In this study several threshold value are investigated and the best one is chosen. It is nice mentioning that different input data with different characteristics may require different *TH* value. In other words, observed data from other wind parks may require different value of threshold as these data sets significantly vary in the number of trivial and redundant feature.

Note that the most relevant attributes to output feature might not lead to the results because it may still contain redundant data. The redundancy of data in the model building phase, in fact, can have an adverse impact on the prediction performance as well as computational cost. Indeed, the selected *m* best attributes may not lead to highest accuracy which can be obtained with the best *m* features. Therefore, the main focus of the second stage of the proposed algorithm is to detect the attributes which are strongly correlated to other attributes and then remove them.

To do so, in this step, a three-layer feedforward neural network is employed as illustrated in Figure 4.2. The network is trained in such a way that redundant features have the associated weights with the lower value. On the other hand, the prominence of features is recognized by the magnitude of their connections from the input layer to hidden layer and the hidden layer to output layer. The weights with the lower value can be eliminated due to the insignificant impact on the accuracy of the network. After deletion of the weights with small magnitude, the accuracy of the network remains markedly preserved and if it reduced, it can be recovered by retraining the network.

Generally, the error function measured during the training process is defined as follows:

$$S_p = \frac{1}{n}\sum\nolimits_{i=1}^{n} \left(t_p(i) - y_p(i)\right)^2 \tag{4.10}$$

where n is the number of samples, p in the number of output y is the actual value and t is the network output which is expressed as follows:

$$y_p(i) = \zeta\left(\sum_{m=1}^{h}\left(\psi\left(\sum_{l=1}^{V} w_l^m X_l(i)\right) * V_p^m\right)\right) \tag{4.11}$$

where X is a $V * n$ matrix, h is the number of hidden units, V is the number of attributes selected in the first stage, w_l^m is the weight connecting from l-th attribute to m-th hidden unit, and v_o^m is the weight connecting from m-th hidden unit to network output. ψ is the sigmoid activation function for the hidden layer and ζ is tangent hyperbolic transfer function for the output layer which are defined respectively as follows:

$$\psi(y) = \frac{1}{1 + \exp(-y)} \tag{4.12}$$

$$\zeta(y) = \frac{\exp(y) - \exp(-y)}{\exp(y) + \exp(-y)} \tag{4.13}$$

In order to detect unnecessary attributes, a penalty function is added to Eq. (4.10) which is expressed as:

$$\begin{aligned} P(w) = \alpha_1 \sum_{m=1}^{h}\left(\sum_{l=1}^{V} \frac{\varsigma(w_l^m)^2}{1 + \varsigma(w_l^m)^2} + \sum_{o=1}^{p} \frac{\varsigma(v_o^m)^2}{1 + \varsigma(v_o^m)^2}\right) \\ + \alpha_2 \sum_{m=1}^{h}\left(\sum_{l=1}^{V}(w_l^m)^2 + \sum_{o=1}^{p}(v_o^m)^2\right) \end{aligned} \tag{4.14}$$

where α_1, α_2 and ς are coefficients that control the influence of the penalty term.

As the key concept of this algorithm, the performance of the trained network $\mathbb{N}$ is observed based on the received input features from the first stage, $XS = \{X_1, \ldots, X_V\}, XS \subset X, V < (\mathrm{NL_{WS}} + \mathrm{NL_T} + \mathrm{NL_{WD}} + \mathrm{NL_H} + \mathrm{NL_{WP}})$. In order to build new models the number of features sequentially decreased. Supposing that $k = \{1,2, \ldots, V\}$, the performance of the network $\mathbb{N}_k$ is observed receiving k less feature compared to the XS and then the algorithm, at the end makes decision if more features can be eliminated. The prime steps of this method are shown in Figure 4.3 and is expounded in the following.

1. Given input vector $XS = \{X_1, \ldots, X_V\}$, $XS \subset X$ with the size of $V * n$ is separated into two data set: training set, S_T and testing set S_C. The network $\mathbb{N}$ is trained to minimize Eqs. (4.10) and (4.14). It also computes the accuracy of the training set R_T, the testing set R_C and also the maximum acceptable decrease (ΔR) in network accuracy using set S_C. It should be noted that in the first procedure of training, equal value of α_1 and α_2 are set for the weights from input layer to hidden layer.
2. Suppose features $XS - XS\{1, \ldots, k\} = XS\{k + 1, \ldots, V\}$ are input features of the network $\mathbb{N}_k$, i.e, $\mathbb{N}_k$ does not include the first k attributes. For instance, $XS\{4,5, \ldots, V\}$ are input features of network $\mathbb{N}_3$. The network $\mathbb{N}_k$ is trained while the connection from attribute $XS(k)$ to hidden layer is set to zero and all weights from other attributes are set equal to the weights of network $\mathbb{N}$. The accuracy of training and testing set for all k is measured and called R_T^k and R_C^k respectively. Based on the research conducted in (Setiono & Liu, 1997) the maximum acceptable decrease (ΔR) is set to 3%.
3. Rank networks $\mathbb{N}_k$ based on testing set as $R_C^1 \geq R_C^2 \geq \cdots \geq R_C^V$. Then, compute the average of this rate R_C^{avg}.

4. Algorithm updates the penalty parameter of attributes. If the accuracy of the network $\mathbb{N}_k$ denoted by R_C^k is smaller than R_C^{avg}, only the weights from attribute $XS(k)$ are multiplied by 1.1. In fact, the expectation is that with the larger penalty parameters, network $\mathbb{N}_k$will produce a smaller magnitude for the weights connected to $XS(k)$. On the contrary, if R_C^k is higher than R_C^{avg} all network connections from input $XS(k)$ are divided by 1.1. This allows salient inputs having connections with higher value in magnitude after network are retrained. On the other hand, algorithm removes network connections having a small magnitude representing unimportant attributes.

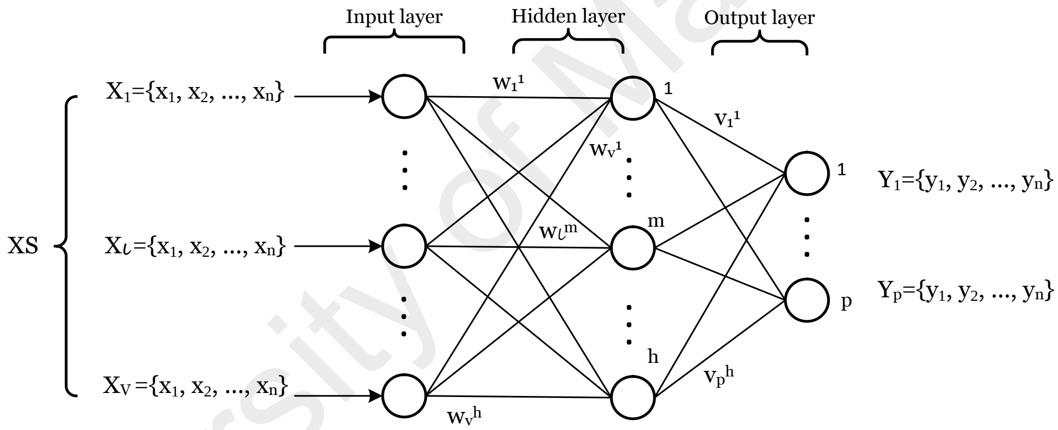


Figure 4.2: The second stage of developed FS algorithm based on NN

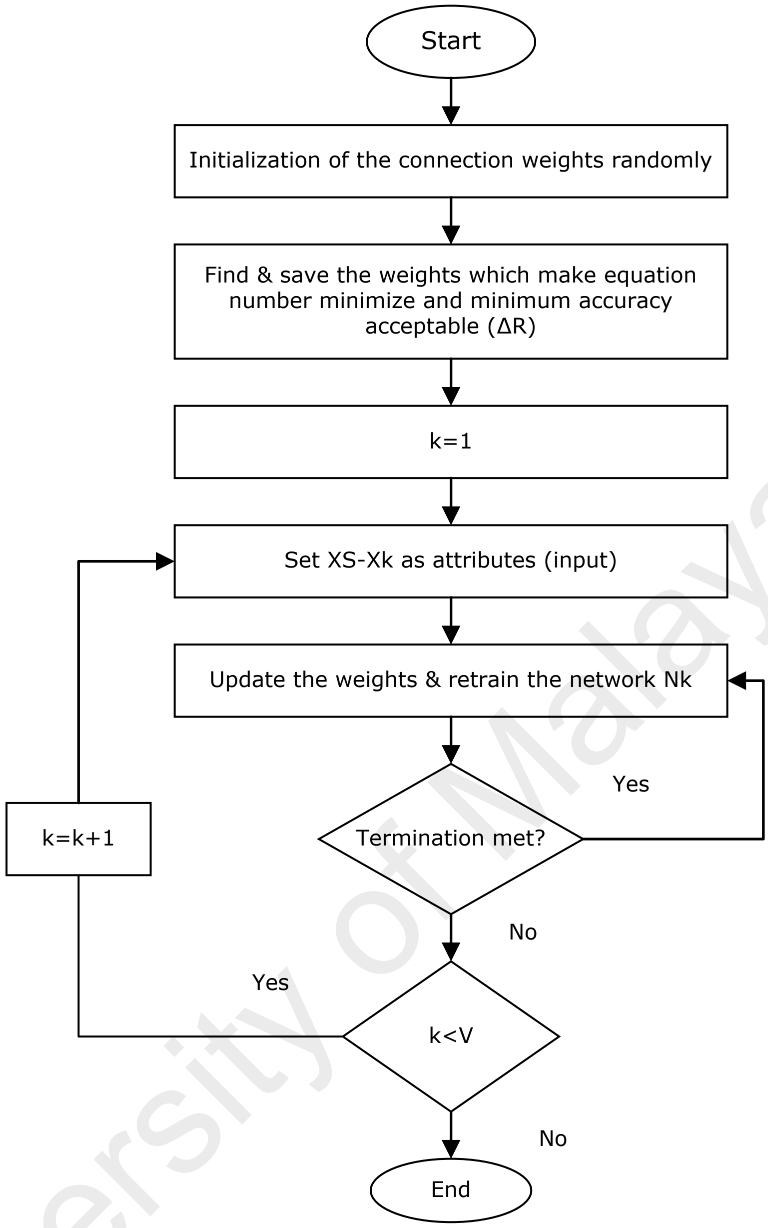


Figure 4.3: Flowchart for the second stage of the developed FS method

It must be noted that the initial setting for the $\alpha_1$and α_2 may not be generalised to all problems. According to this research and the carried out survey by (Setiono & Liu, 1997), the recommended initial settings for the α_1, α_2, and ς are 10^{-1}, 10^{-4}, and 10 respectively.

4.5 Direct Wind Power Forecasting

Unlike parametric and statistical models, the data-driven approaches do not involve equations. Indeed, the primary focus of these approaches is to discover a pattern between wind direction, wind speed, and temperature. In the current section, adaptive neuro-fuzzy inference system (Schlechtingen et al., 2013) and five other data mining algorithms namely, *k*-nearest neighbor (Yesilbudak, Sagiroglu, & Colak, 2013), M5Rules (Lydia et al., 2013), random forest (Lahouar & Slama, 2015), support vector machine (L. Yang, He, Zhang, & Vittal, 2015) and multilayer perceptron (Velo, López, & Maseda, 2014) are used to forecast the produced wind power. Data-driven methods are explained in detail in the sections below.

4.5.1 k-Nearest Neighbors Algorithm (k-NN)

The main idea of k-NN is that the k nearest neighbors of the new sample are chosen from the training dataset to forecast the output of the sample. In other words, the prediction of the new point can be obtained by the average of the k nearest neighbors' values. K-NN algorithm is based on only memory and does not use any model to fit. The process of the wind power prediction using k-NN algorithm can be divided into three steps: calculation of the pre-defined distances between the training and testing example including wind speed, wind direction, temperature; choosing k nearest neighbors from the training dataset as per calculated distance; and prediction of the wind power using a weighted average technique.

Firstly, a distance metric is required to measure the closeness of any two observations. Euclidean distance though is often used to characterize the similarity of the data points, for further evaluation, the Manhattan distance and the Mahalanobis distance are also applied in this research which can be expressed as follows:

Euclidean distance metric
$$D[X,Y] = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{1/2} \tag{4.15}$$

Manhattan distance metric
$$D[X,Y] = \sum_{i=1}^{n}|x_i - y_i| \tag{4.16}$$

Minkowski distance metric $$D[X,Y] = \left[\sum_{i=1}^{n}(x_i - y_i)^m\right]^{1/m} \quad (4.17)$$

where X and Y are two observations instances from the training and testing datasets, respectively; x_i and y_i are the input variables selected by the developed feature selection and n is the number of input variables. According to Eq. (4.17), it can be clearly seen that Euclidean and Manhattan distance metrics are special cases of Minkowski where the parameter m is set to 2 and 1 respectively. Additionally, parameter m is usually set to 3 in Minkowski distance metric.

Secondly, the observations with the k smallest distance, representing the greater similarity, are selected as the k nearest neighbor. Assuming, $X_1, X_2, \dots, X_k$ are the k nearest neighbors, $p_1, p_2, \dots, p_k$ are the corresponding wind power, and $d_1 \le d_2 \le \cdots \le d_k$ are the distance from the training sample X_k to the test sample Y which is sorted ascendingly, then the predicted wind power is derived as follows:

$$p_e = \sum_{i=1}^{k} w_k p_k = \frac{\sum_{i=1}^{k} \exp(-d_k) p_k}{\sum_{i=1}^{k} \exp(-d_k)} \quad (4.18)$$

The parameter w shows that the importance of each observation is not considered equally. In fact, the closer the neighbor is the greater weight it obtains. Consequently, the higher influence its corresponding wind power has on forecasted wind power, p_e.

4.5.2 M5-Rules

The idea behind this algorithm is to recursively partition the data space and fitting a simple prediction model within each partition. This algorithm technically generates a set of rules from M5 model tree. In this algorithm, firstly a tree learner is employed to training dataset and after learning the pruning technique is used. Usually pruning is required to reduce the size of decision tree results in improvement in the predictive accuracy and the reduction of overfitting. In the next step, the best node is chosen and converted to a rule. The observations covered by the rule are removed from the dataset. The process is applied recursively to the remaining observations and terminates when all samples are covered by one or more rules. In this research, picking up the most informative node is based on the percentage of root mean square (RMS) error. In this case, small values of % RMS represents that the model at a leaf is performing better than simply predicting the mean of the class values. RMS value is given by:

$$\%\text{RMS} = \frac{\left[\sum_{i=1}^{n_r} \frac{\left(p_a(i) - p_e(i)\right)^2}{n_r}\right]^{1/2}}{\left[\sum_{i=1}^{n_r} \frac{(p_a(i) - \overline{p_a})^2}{n}\right]^{1/2}} \tag{4.19}$$

where $p_a$is the actual value, p_e is the predicted value by linear model, n_r is the number of observations covered by the leaf, $\overline{p_a}$ is the average of actual values , and n is the total number of samples. Figure 4.4 illustrates a simple decision tree with two inputs x_1 and x_2 and based on 5 nodes.

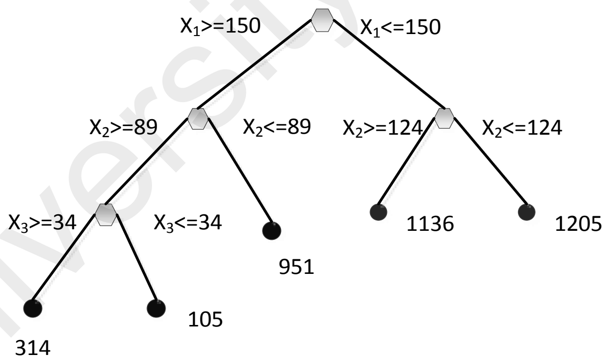


Figure 4.4: M5-rules regression tree

4.5.3 Random Forest (RF)

Random forest is a kind of nonparametric machine learning that includes a multitude of decision trees and outputs the result, which is the mean prediction of the individual trees. Generally, a tree is a set of nodes and branches organized in a hierarchy with no loops. RF is broadly similar to bootstrap aggregation (bagging), but additionally, a randomized subset of predictors is selected for each split of each tree. In fact, random selection of features reduces the correlation between trees, which results in further improvement in prediction accuracy. In general, the robustness and the immunity to outliers are two advantages of RF. The procedures of RF approach are explained as follows:

- Let the number of instances be N and the number of features be n.
- The number of features at a node of the decision tree is determined to be m ($m < n$).
- The following steps are repeated for each decision tree:
- A subset of training data is set with a replacement that represents the N instances and the rest of the data is used to measure the error of the tree.
- The following step is repeated for each of node of this tree

In order to determine the decision at this node and calculate the best split accordingly, m number of features are selected randomly. Tree pruning is prohibited.

4.5.4 Support Vector Machine (SVM)

SVM is a kind of machine learning algorithm which attempts to discover a nonlinear map from input space to output space and map the data in high-dimensional feature space through the map. Then, the following equation is applied to the given training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ to perform a linear regression in this feature space.

$$f(x) = w_1 x_1 + \cdots + w_n x_n + b = \mathbf{W}^T \mathbf{X} \tag{4.20}$$

where **W** and b are the weight vector and the bias respectively which can be estimated by minimizing by the following constrained optimization problem:

$$\text{Minimize} \quad \frac{1}{2}\|\mathbf{W}\| + C\sum_{i=1}^{n} \xi_i + \xi_i^*$$

$$\begin{cases} (\mathbf{W}^T x_i + b) - y_i \leq \epsilon + \xi_i \\ y_i - (\mathbf{W}^T x_i + b) \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{4.21}$$

where ξ and ξ_i^* represent the slack variables, C and ϵ are the regularization parameter and tolerance threshold respectively. By using Lagrange multipliers, Eq. (4.21) can be reformulated as:

$$\text{Maximize} \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) - \epsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i)$$
$$-\frac{1}{2}\sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \langle x_i, x_j \rangle \tag{4.22}$$

$$\text{Subject to} \begin{cases} \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0 \\ \alpha_i^*, \alpha_i \geq 0 \end{cases} \tag{4.23}$$

where $\alpha_i^*$and α_j are the Lagrange multipliers and $\langle ., . \rangle$ denotes the inner product which can be substituted with Kernel function as follows to avoid any curse of dimensionality.

$$f(x) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) K(x, x_i) + b \tag{4.24}$$

Finally, the coefficient parameters of the equation number are computed as:

$$\langle w, x \rangle = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) K(x, x_i) \tag{4.25}$$

$$b = -\frac{1}{2}\sum_{i=1}^{n} (\alpha_i^* - \alpha_i)\big(K(x_i, x_r) + K(x_i, x_s)\big) \tag{4.26}$$

where x_r and x_s represents the identified support vectors. Noted that a commonly applied kernel function is a radial basis function (RBF) which is defined as:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{4.27}$$

where x is the testing dataset, x_i is a vector parameter, and σ is the kernel adjustable parameter.

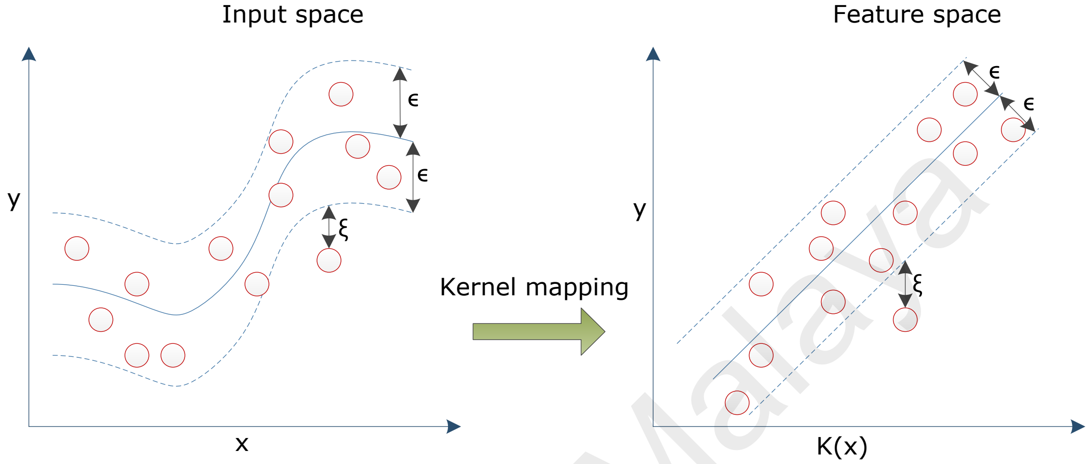


Figure 4.5: Graphic illustration of kernel mapping

4.5.5 Multilayer Perceptron (MLP)

MLP is a supervised algorithm, with one or more layers between input and output layer, mapping input dataset onto a set of appropriate outputs. MLP is a kind of feedforward neural network as data flows in one direction from input to output. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then sending the output through some nonlinear transfer functions which can be mathematically expressed as:

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi(\boldsymbol{W}^T \boldsymbol{X} + b) \tag{4.28}$$

where $\boldsymbol{W}$ is the vector of weights, b is the bias, φ is transfer function, $\boldsymbol{X}$ is vector of inputs, and y is the output.

4.5.6 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a class of adaptive multilayer feedforward networks, applied to nonlinear forecasting where past samples are employed to predict the future sample. ANFIS combines the self-learning ability of NN and the linguistic expression function of fuzzy inference. As shown in Figure 4.6, ANFIS comprises five layers.

The first layer consists of input variables membership functions (MFs), input 1 and input 2. In layer 1, every node i is an adaptive node with node function:

$$O_i^1 = \gamma A_i(x) \qquad i = 1,2 \tag{4.29}$$

or

$$O_i^1 = \gamma B_{i-2}(y) \qquad i = 1,2 \tag{4.30}$$

where O_i^1 is the output of the ith node in layer 1, x and y are inputs, A_i and B_{i-2} are the linguistic labels such as ‘large’ or ‘small’. γ is the membership function for A and B which is usually the generalized bell-shaped function expressed as follows:

$$\gamma A_i(x) = \frac{1}{1 + \left|\frac{x - r_i}{p_i}\right|^{2q_i}} \tag{4.31}$$

where $\{r_i, p_i, q_i\}$ is the variable set. The bell-shaped function varies as the values of the variables change, hence manifesting various types of membership functions for fuzzy set A. Variables in the first layer are called premise variables.

In the second layer, each node performs connective operation ‘‘AND’’ within the rule antecedent to determine the corresponding firing strength, w_i. The product of this layer is written as follows:

$$O_i^2 = w_i = \gamma A_i(x)\gamma B_i(y), \quad i = 1,2 \tag{4.32}$$

Layer 3 generates the normalized firing strength expressed as follows:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \qquad i = 1,2 \tag{4.33}$$

Layer 4 is a defuzzification layer wherein for each node the contribution of the *i*th rule to the overall output is computed as follows:

$$O_i^4 = \overline{w}_i z_i = \overline{w}_i (a_i(x) + b_i(y) + c_i), \quad i = 1,2 \tag{4.34}$$

where $\{a_i, b_i, c_i\}$ are the constant parameters.

The final output as a summation of all inputs is computed as equation number by single node $\sum$ in the last layer. The fifth layer is not adaptive and transforms the fuzzy classification results into a crisp output. The output represents the generated output power.

$$O_i^5 = \sum_i \overline{w}_i z_i = \frac{\sum_i w_i z_i}{w_i} \tag{4.35}$$

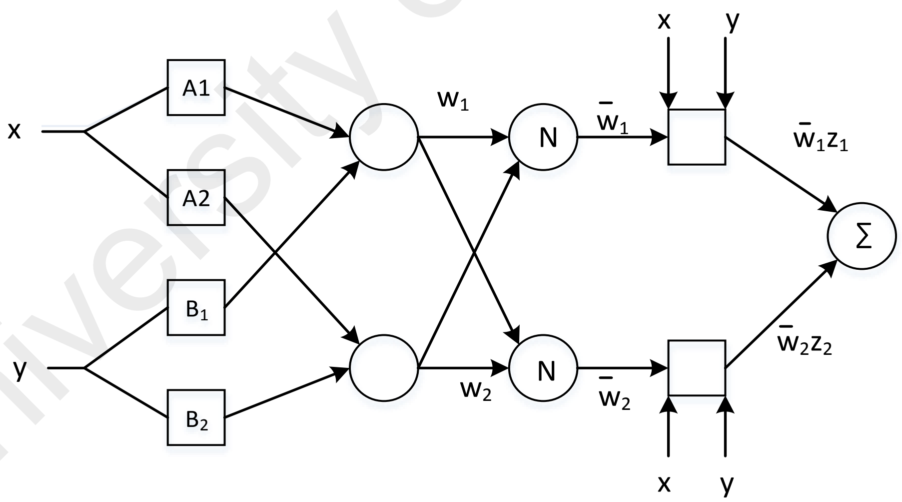


Figure 4.6: ANFIS structure with two inputs, two rules, and one output

4.6 Real Data

The same real data in chapter 3, 5-min wind direction, wind speed, temperature, and output power, covering the period of 12 months are employed in this chapter. The observations are divided into two sets, data set 1 and data set 2. The former comprises of 86206 instances which are employed to develop a model for wind power and wind speed prediction. Data set 2 includes 8061 observations and covers four weeks test data which are randomly selected, corresponding to the four seasons in a year: the first week of November (fall), the third week of May (spring), the third week of August (summer), and the fourth week of February (winter).

4.7 Prediction Evaluation Indicators

To examine the performance of applied models several assessment criteria are required. In this chapter three forecasting evaluation indicators are employed: mean absolute percentage error (MAPE) which was expressed in chapter 3 as Eq. (3.68); mean absolute error (MAE) and standard deviation error (SDE) as given by:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_e - y_a| \tag{4.36}$$

$$\mathrm{SDE\ of\ MAE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(|y_e - y_a| - \frac{1}{n}\sum_{i=1}^{n}|y_e - y_a|\right)^2\right]^{1/2} \tag{4.37}$$

where y_e is the estimated power, y_a is the actual power and n is the number of samples forecasted.

4.8 Simulation Results

This section includes the simulated results obtained by indirect and direct wind power prediction methods.

4.8.1 Indirect Wind Power Prediction Results

In this research, the auto-correlation function (ACF) is conducted as a visual inspection to reaffirm the stationary of the data applied in ARMA. A fast upward trend in ACF plot presents the stationary of the data, whereas the gradual downtrend shows the data is non-stationary. After the determination of ARMA (p,q) structure, the parameters of the model is estimated through maximum likelihood estimation using the Kalman filter in conjunction with Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. After close observation, ARMA (2,1) is selected for wind speed forecasting. To calculate the coefficient of DES model, Eq. (4.38) is minimized by BSA. In this paper for BSA, the maximum number of iterations is 2000, population size is 30, *mixrate*=0.95, *F*=3.*rndn*, whereby stopping criteria is based on the maximum number of iteration. Noted that *rndn* is the normal distribution with mean equal to 0 and standard deviation equal to 1.

$$\text{Objective function: } \min \sum_{t=1}^{n} (y_{t+1} - \hat{y}_{t+1})^2$$

$$\hat{y}_{t+1} = (1 - \eta - \lambda)(S'_{t-1} + b_{t-1}) + \eta y_t + \lambda S'_t + b_{t-1} \tag{4.38}$$

$$0 < \eta, \lambda < 1$$

Table 4.1 presents the performance of DES and ARMA, and persistence5-min data in winter, spring, summer, and fall. According to comparative results, it can be observed that the MAPE, MAE and the SDE values obtained with the DES model are considerably lower than those obtained with the other two models. The value of unknown parameters of DES (η, λ) using BSA is illustrated in Table 4.2. For the further evaluation, the prediction of wind speed over 15-, 30-, and 60-min ahead the same procedure is performed. In this context, to make 15-, 30-, and 60-min ahead forecasting, three, six, and twelve consecutive data in section 2.1 should be averaged. Table 4.3 denotes the performance of DES model over 5-, 15-, 30-, and 60-min ahead prediction based on four test week data. It can be deduced from Table 4.3 that statistical error measures increase by prediction interval, with the exception of summer and spring over 15- and 30-min time interval. Note that, the ARMA and persistence models also were evaluated for different prediction horizon, however, the results are not shown here because the models displayed similar trends to those of the real data. Figure 4.7 shows the actual and the forecasted wind speed using time series model over 5-min averaged data.

Table 4.1: Comparison of wind speed forecasting methods based on data set 2 (5-min)

Model	Error	Test week				Average
		Winter	Spring	Summer	Fall	
DES	MAPE	0.34	0.27	0.68	0.45	0.43
	MAE	0.03	0.03	0.09	0.02	0.04
	SDE	0.06	0.06	0.10	0.02	0.06
ARMA	MAPE	4.7	6.12	7.02	8.25	6.52
	MAE	0.50	0.78	0.93	0.33	0.63
	SDE	0.88	1.32	1.56	0.57	1.08
Persistence	MAPE	6.93	6.85	7.17	8.34	7.32
	MAE	0.46	0.85	1.00	0.48	0.69
	SDE	0.87	1.46	1.68	0.92	1.23

Table 4.2: Estimated parameters value of DES model based on dataset 2 (5- to 60-min)

Time Horizon	Parameter	Test week			
		Winter	Spring	Summer	Fall
5-min	η	0.998	0.996	0.999	0.998
	λ	0.014	0.011	0.002	0.003
15-min	η	0.998	0.998	0.995	0.997
	λ	0.030	0.003	0.016	0.015
30-min	η	0.996	0.998	0.994	0.997
	λ	0.025	0.005	0.020	0.024
60-min	η	0.992	0.996	0.993	1.00
	λ	0.032	0.015	0.078	0.067

Table 4.3: Statistical error measures of DES over 5- to 60-min ahead prediction

Error	Test week				Average
	Winter	Spring	Summer	Fall	
MAPE (%)	0.34	0.27	0.68	0.45	0.4350
MAE	0.03	0.03	0.09	0.02	0.0425
SDE	0.06	0.06	0.10	0.02	0.0600
MAPE (%)	0.78	0.20	0.34	0.59	0.4775
MAE	0.08	0.02	0.04	0.02	0.0400
SDE	0.13	0.02	0.08	0.04	0.0675
MAPE (%)	1.37	0.26	0.51	2.16	1.0750
MAE	0.14	0.33	0.07	0.09	0.1575
SDE	0.33	0.02	0.10	0.10	0.1375
MAPE (%)	2.17	0.46	1.59	3.00	1.5112
MAE	0.23	0.06	0.21	0.12	0.1330
SDE	0.51	0.06	0.38	0.15	0.2317

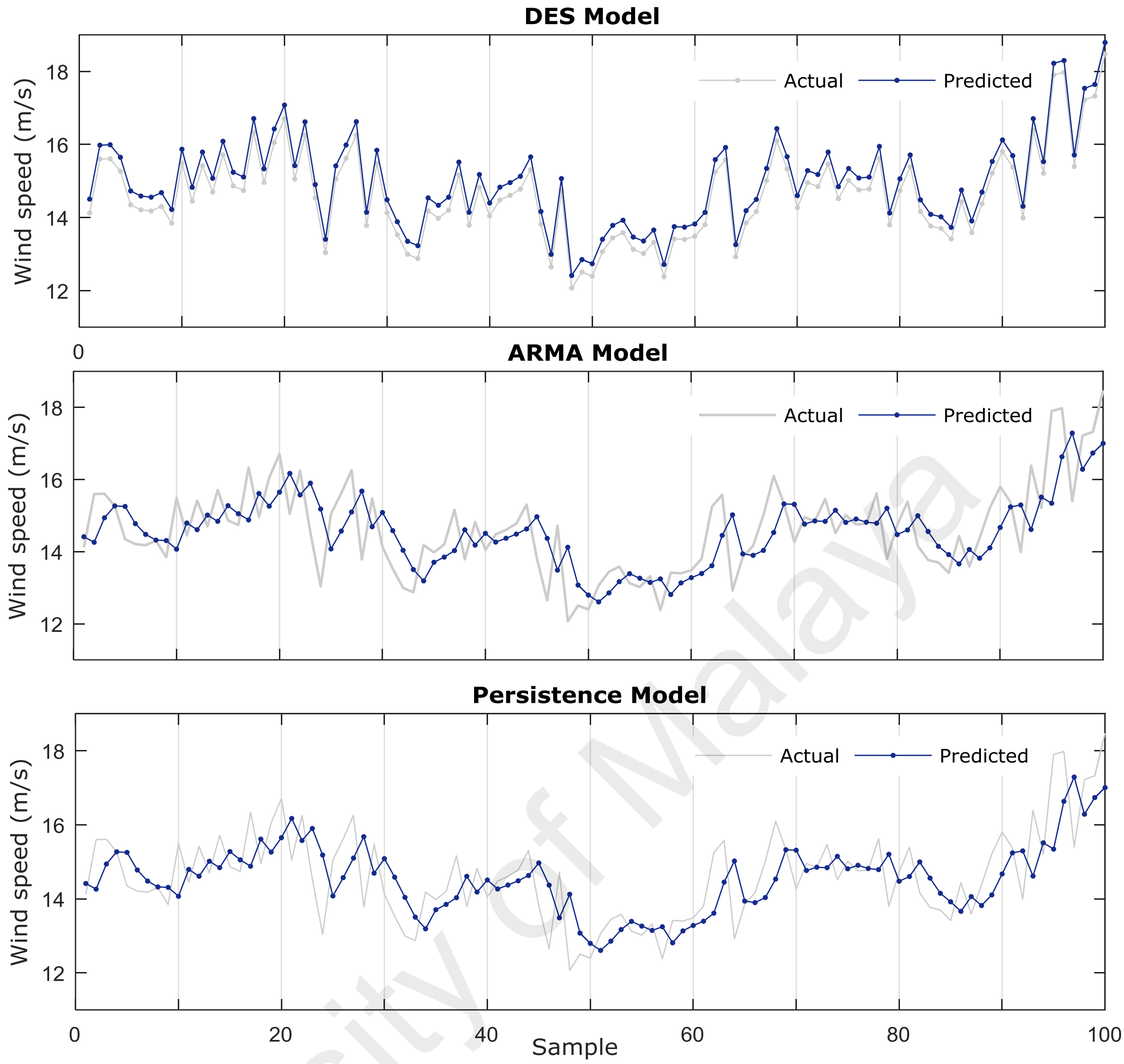


Figure 4.7: The performance of wind speed prediction methods over the first 100 samples of the test data in August

After prediction of wind speed, MHTan wind turbine power curve modeling (Eq. (3.1), which was described in details in chapter 3, can be used to estimate the future wind power. Table 4.4 summarizes the forecasted power by MHTan based on four seasons over four time interval. Figure 4.8 shows the performance of MHTan based on the last 100 samples in the test week in May.

Table 4.4: 5- to 60-min ahead prediction results of indirect prediction using MHTAn

Error	Time	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	5-min	1.83	1.60	1.11	9.09	3.41
	15-min	2.14	1.40	1.05	8.11	3.17
	30-min	2.31	1.32	1.00	8.25	3.22
	60-min	2.64	1.39	1.68	8.78	3.62
MAE	5-min	18.23	18.83	14.33	20.44	17.96
	15-min	21.33	16.42	13.53	18.40	17.42
	30-min	21.95	15.49	12.87	18.49	17.20
	60-min	26.32	16.30	21.67	19.66	20.98
SDE	5-min	37.40	35.36	24.81	33.69	32.81
	15-min	38.34	30.53	32.86	29.38	32.78
	30-min	42.31	29.83	30.75	26.89	32.44
	60-min	53.05	30.37	49.17	28.25	40.21

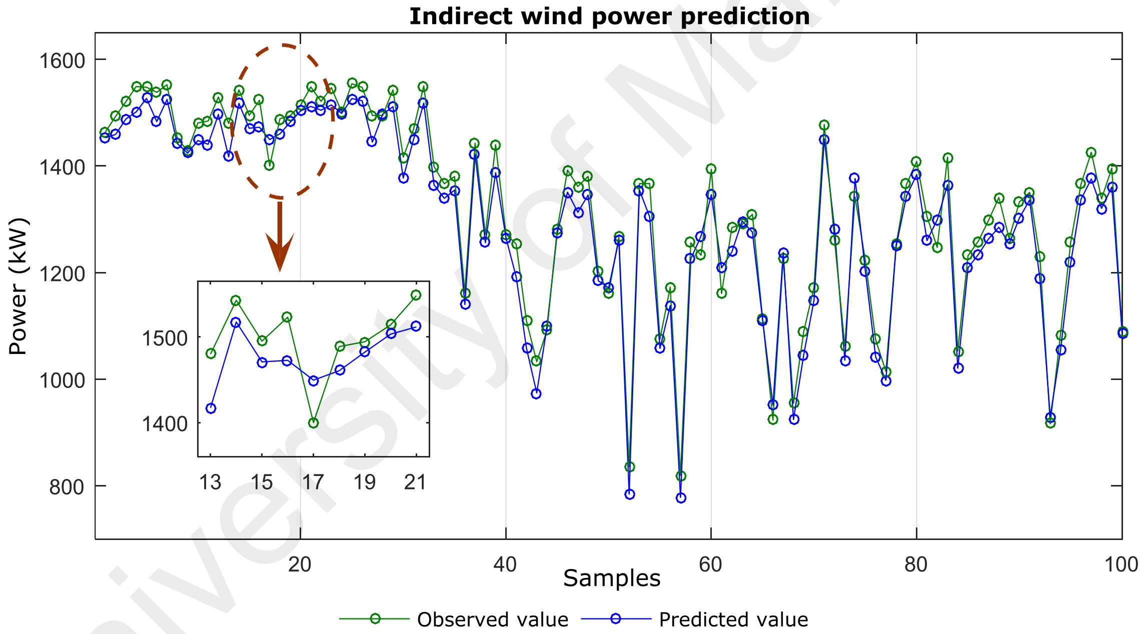


Figure 4.8: Comparison of the actual and forecasted wind power using MHTan based on the last 100 samples in the test week in May

4.8.2 Direct Wind Power Forecasting Results

The most important inputs selected by the developed FS technique in section 4.4 are employed to assess the performance of the mentioned algorithms in section 4.4. In this research, 82 attributes, including 20 lagged wind direction, temperature, wind speed, output power and the current temperature and wind speed are considered which cover almost all the necessary and informative data. It is nice mentioning that due to unavailability of the humidity data, it is not considered in this study. Considering TH=0.64, 33 out of 82 inputs as the relevant features are selected in the first stage. In the next stage, 16 attributes of the data selected in the first stage, are detected as redundant feature and then are removed. Eventually, only 17 data with the maximum relevancy and the minimum redundancy are used as inputs for the six before-mentioned algorithms. Table 4.5 denotes the selected features.

To validate the performance of the developed FS technique, it is compared with other methods such as correlation analysis (CA) (Hong et al., 2010), principal component analysis (PCA) (Kong, Liu, Shi, & Lee, 2015) and relief feature selection techniques (Koutanaei, Sajedi, & Khanbabaei, 2015). CA only discover linear relationship between two variables. PCA linearly transforms the original inputs into new uncorrelated features and relief is a feature weight-based algorithm inspired by the instance-based learning algorithm. The main drawback of the above-mentioned FS methods is that, they are not able to detect the redundancy of the data. This issue, however, is improved by the second stage of the developed FS technique. The performance of the FS techniques based on different test weeks are shown Table 4.6. In order to draw a better comparison, the average value of the error for each individual technique is presented as well. The comparative results clearly prove the superiority of the developed FS method compared to the other methods. Wind speed is vitally important data, considering that the better rank represents the more significant influence of the corresponding feature on wind power.

Table 4.5: Selected features by developed FS method based on the test week in August

Selected Attributes	Rank	Selected Attributes	Rank	Selected Attributes	Rank
$WP(t-1)$	1	$T(t-1)$	7	$WS(t-12)$	13
$WS(t)$	2	$WS(t-4)$	8	$WD(t-8)$	14
$WS(t-1)$	3	$WP(t-5)$	9	$WP(t-17)$	15
$WP(t-3)$	4	$WD(t-2)$	10	$WS(t-18)$	16
$WP(t-8)$	5	$WS(t-9)$	11	$T(t-10)$	17
$WS(t-2)$	6	$T(t-5)$	12		

Table 4.6: Comparison of different FS technique based on data set 2

Season	PCA			CA			Relief			MI+NN		
	MAPE	MAE	SDE	MAPE	MAE	SDE	MAPE	MAE	SDE	MAPE	MAE	SDE
Winter	0.88	8.83	17.28	0.96	8.81	18.05	0.86	8.61	17.04	0.76	7.59	16.09
Spring	1.52	17.80	31.39	1.38	16.14	29.02	1.36	15.96	29.77	1.34	15.76	29.04
Summer	1.06	13.67	24.82	0.91	11.74	23.50	0.94	12.18	24.07	0.83	10.79	22.09
Fall	3.47	7.86	13.91	3.49	7.90	14.46	3.04	6.88	13.28	2.96	6.66	12.76
Average	1.73	12.04	21.85	1.69	11.15	21.26	1.55	10.90	21.04	1.47	10.20	19.99

The performance of the ANFIS and data mining algorithms is evaluated based on MAPE, DAE, and SDE as indicated in Table 4.7. To make a fair comparison between direct and indirect wind power forecasting, the data used in section 4.8.1 is employed in this section too. The table indicates that k-NN and MLP models yield satisfactory and quiet similar accuracy. It also discloses that ANFIS achieves the best results while the performance of random forest is disappointing. ANFIS, therefore, is selected to forecast wind power over longer horizon up to 60-min.

Table 4.7: MAPE, MAE, and SDE results of different direct methods in wind power forecasting based on data set 2

Error	Algorithm	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	k-NN (k=50)	0.81	1.43	0.86	3.12	1.55
	M5Rules	1.11	1.55	0.84	3.32	1.70
	Random forest	1.00	1.69	1.06	4.88	2.15
	SVM	0.87	1.41	1.13	3.19	1.65
	MLP	0.82	1.43	0.87	3.13	1.56
	ANFIS	0.76	1.34	0.83	2.96	1.47
MAE	k-NN (k=50)	8.07	16.87	11.12	7.03	10.77
	M5Rules	11.15	18.28	10.91	7.48	11.95
	Random forest	10.02	19.93	13.68	10.97	13.65
	SVM	8.70	16.58	14.61	7.18	11.76
	MLP	8.19	16.78	11.27	7.03	10.81
	ANFIS	7.59	15.76	10.79	6.66	10.20
SDE	k-NN (k=50)	17.99	32.77	20.98	13.00	21.18
	M5Rules	21.89	34.40	21.12	13.66	22.76
	Random forest	20.93	37.47	26.26	18.49	25.78
	SVM	24.86	31.81	23.05	13.16	23.22
	MLP	17.97	32.61	21.79	13.00	21.34
	ANFIS	16.09	29.04	22.09	12.76	19.99

In machine learning viewpoint, selecting the best parameters of NN has been a chronic problem. In other words, there is no definite technique to determine the best structure of NN. Moreover, a particular structure which is performing well in a particular problem may not guarantee the same performance in other problems. Hence, in this research, to determine the best parameters of NN, fifteen structures are tested. For example, the performance of MLP is observed over a different type of transfer functions, the number of nodes, number of hidden layers, etc. Finally, MLP was developed with the following structure: two hidden layers, back propagation learning algorithm, hyperbolic tangent sigmoid transfer function, six nodes in the first hidden layer and three nodes in the second hidden layer, and the number of iterations of 1000. Similarly for SVM, a set of value on regularization parameter, C, at $\{2^0, 2^3, 2^7, 2^{10}\}$ and RBF kernel width, σ, at $\{$ 0.5,1.0,1.5,2.0$\}$ were evaluated and then the optimal value of $C = 23$ and $\sigma = 1.0$ was determined by consideration of all possible combinations of parameter values. The number of trees in RF and the minimum number of samples per leaf in M5rules were set to 500 and 10 respectively, whereas ANFIS Sugeno type was set up with 8 cluster centers. The k-NN also disclosed different results over different numbers of k as well as distance metric. Several examinations, therefore, were conducted to select the optimum number of k and the best metric function . Although the same procedure applied to all direct wind power prediction models to obtain their best parameter, for the sake of conciseness, only the examination results of k-NN are tabulated here. It can be noticed from Table 4.8, although k-NN using Euclidean distance shows better accuracy for $k = 100$ and Minkowski for $k = 200$, overall k-NN achieves more accurate results with Manhattan distance compared to Euclidean and Minkowski. Since k-NN model using Manhattan for k =50 outperforms others, its performance is evaluated in other test weeks as given in Table 4.9.

Table 4.8: k-NN performance in wind power forecasting using three different distance metrics based on test week in summer

Nearest neighbor	Euclidean distance			Manhattan distance			Minkowski distance		
	MAPE	MAE	SDE	MAPE	MAE	SDE	MAPE	MAE	SDE
k=50	0.93	11.23	21.24	0.86	11.12	20.98	0.91	11.30	21.31
k=100	0.90	12.45	21.73	0.95	12.31	21.85	1.04	12.67	22.06
k=150	1.18	14.29	25.17	1.10	14.15	25.31	1.24	14.48	26.10
k=200	1.42	16.98	31.25	1.29	16.71	31.14	1.18	16.90	31.09
k=250	1.55	19.28	37.80	1.47	19.05	37.65	1.59	19.10	37.74

Table 4.9: Performance of k-NN using Manhattan distance for different numbers of k

Error	k-NN algorithm	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	k=50	0.81	1.43	0.86	3.12	1.55
	k=100	0.84	1.42	0.95	3.14	1.58
	k=150	0.88	1.41	1.10	3.15	1.63
	k=200	0.94	1.41	1.29	3.15	1.69
	k=250	1.05	1.41	1.47	3.22	1.78
MAE	k=50	8.07	16.87	11.12	7.03	10.77
	k=100	8.44	16.73	12.31	7.07	11.13
	k=150	8.83	16.63	14.29	7.90	11.91
	k=200	9.39	16.60	16.71	7.10	12.45
	k=250	10.46	16.59	19.05	7.26	13.34
SDE	k=50	17.99	32.77	20.98	13.00	21.18
	k=100	18.76	32.46	21.85	13.16	21.55
	k=150	19.60	32.28	25.31	13.16	22.58
	k=200	20.79	32.18	31.14	13.23	24.33
	k=250	23.21	23.05	37.65	13.64	24.38

Table 4.10 indicates the performance of ANFIS model over 5-, 15-, 30-, and 60-min ahead interval. The table reveals that ANFIS algorithm achieves better results in a shorter horizon than a longer horizon of prediction. Additionally, a difference in obtained errors over different test weeks can be clearly noticed. The main reason for that might be because of the dissimilarity of the wind speed distribution and also the adverse effects of weather conditions on the mechanical efficiency of the turbine. The former can be reaffirmed by obtaining the scale parameter (c) and shape parameter (k) value of the actual data. In this context, the probability distribution of wind speed in the winter and fall is Weibull because k is between 0 and 1, while observed $k > 3$ shows that wind speed follows the normal distribution in the spring and summer. Low temperatures have adverse impact on different materials utilised in the fabrication of wind turbies such as composite and steel materials. Steels become more brittle and composite materials will be subjected to a residual stress. Sufficiently high stresses can cause microcracks in the material which result in stiffness and impermeability reduction of the materials. Low temperature can also damage electrical equipment such as yaw

drive motors and transformers as well as the winding can suffer from a thermal shock. Moreover, long exposure to cold can damage gearboxes and hydraulic couplers. The colder temperature becomes, the more viscosity of the lubricant increases and will damage gears because oil cannot freely circulate. Furthermore, seals and rubber parts loose at low temperature. All these may not necessarily cause part's failure but can result in decreased in performance. The ANFIS performance over 5-, 15-, 30- and 60-minute averaged data are shown in Figures 4.9 to 4.12.

Table 4.10: Prediction errors using ANFIS model over 5- to 60- min ahead prediction

Error	Time	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	5-min	0.76	1.34	0.83	2.96	1.47
	15-min	0.71	1.27	0.92	2.59	1.37
	30-min	0.77	1.23	0.89	2.33	1.30
	60-min	1.03	1.26	0.91	2.72	1.48
MAE	5-min	7.59	15.76	10.79	6.66	10.20
	15-min	7.13	14.96	11.94	5.88	9.97
	30-min	7.72	14.51	11.59	5.23	9.76
	60-min	10.29	14.78	11.81	6.10	10.74
SDE	5-min	16.09	29.04	22.09	12.76	19.99
	15-min	14.78	28.82	29.63	11.01	21.06
	30-min	16.32	27.77	27.34	9.79	20.30
	60-min	25.89	26.78	25.93	12.38	22.74

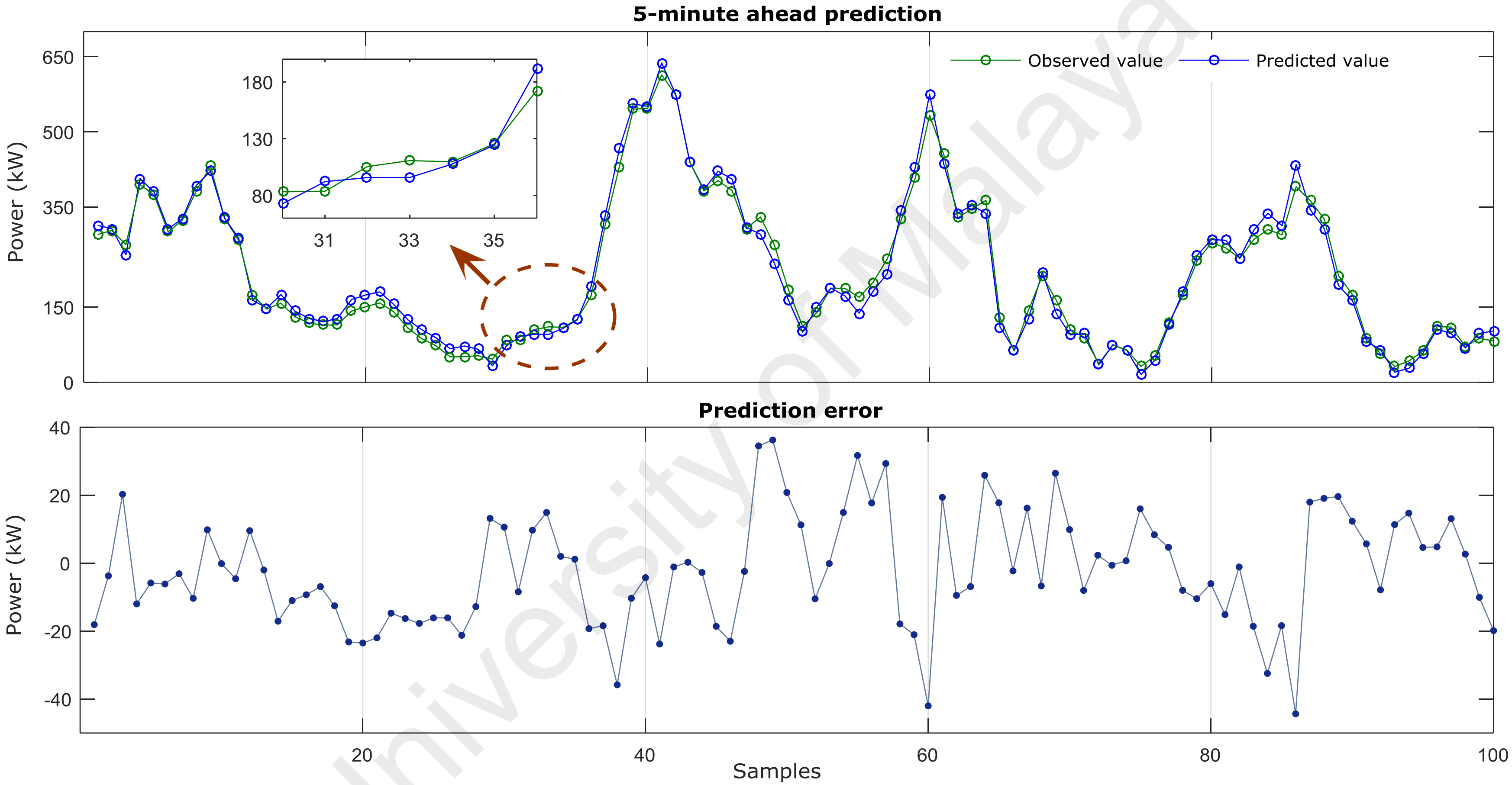


Figure 4.9: Comparison of the actual and forecasted wind power by ANFIS based on 5-min average data of the last 100 samples of the test week in November

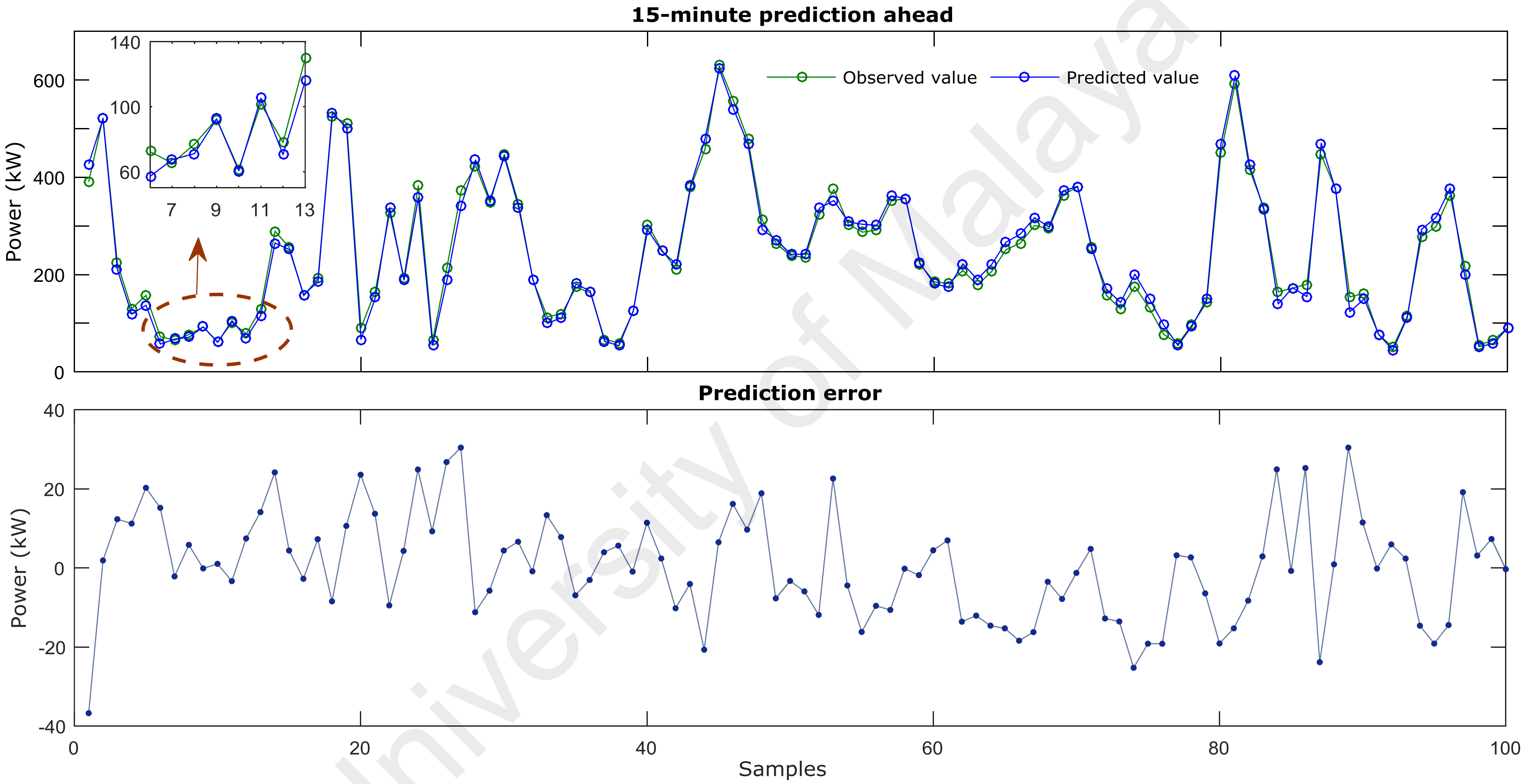


Figure 4.10: Comparison of the actual and forecasted wind power by ANFIS based on 15-min average data of the last 100 samples of the test week in November

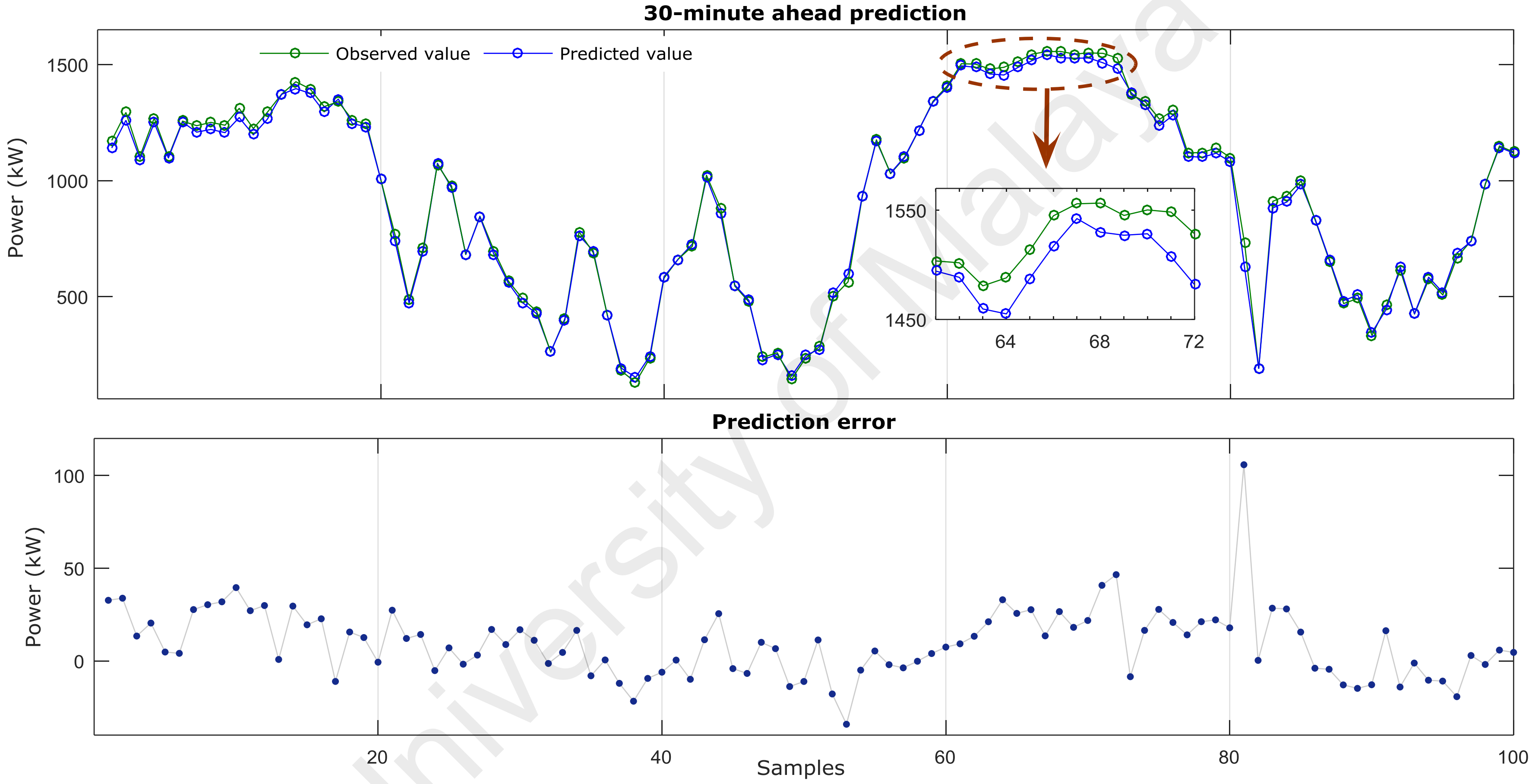


Figure 4.11: Comparison of the actual and forecasted wind power by ANFIS based on 30-min average data of the last 100 samples of the test week in May

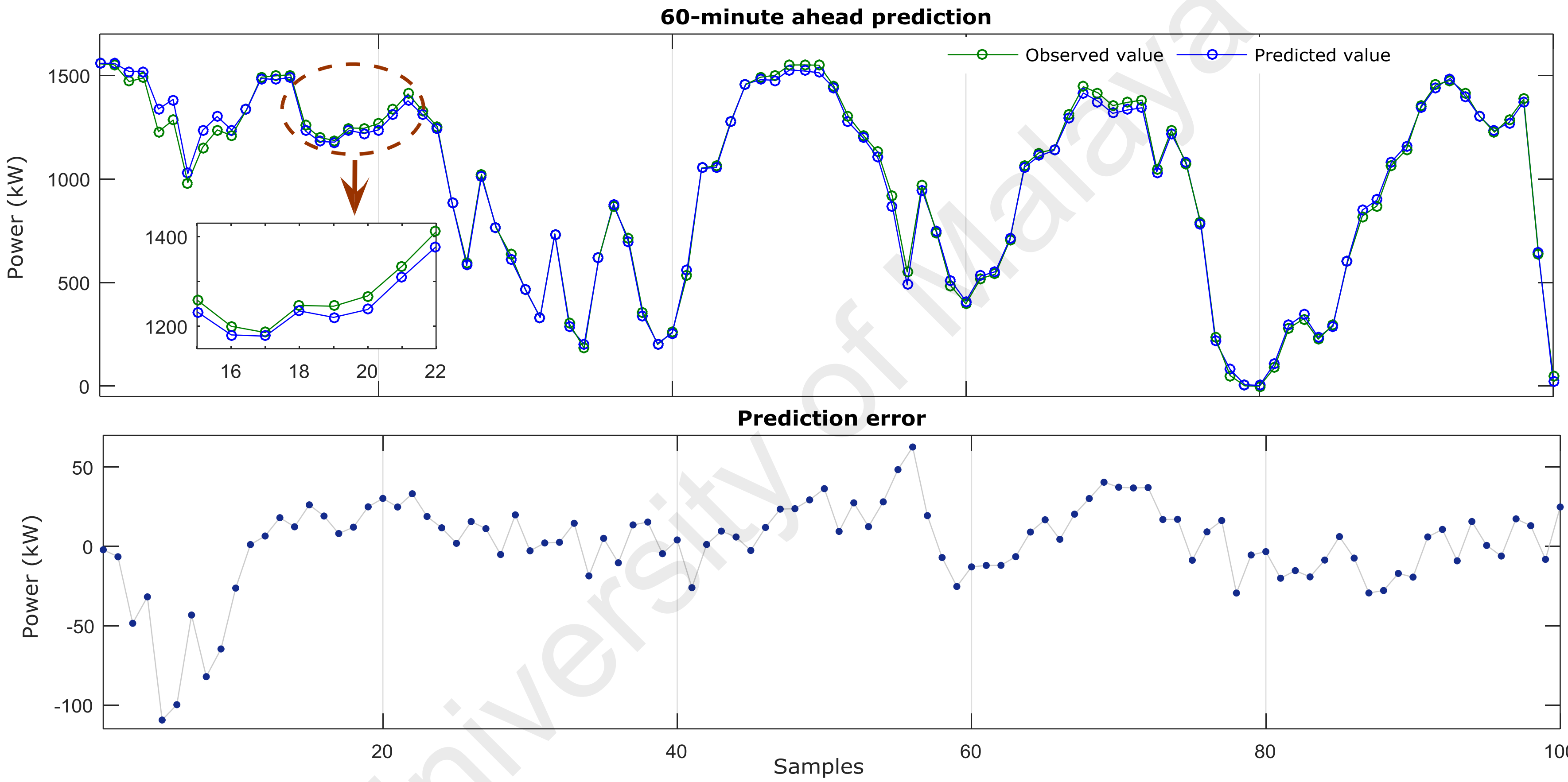


Figure 4.12: Comparison of the actual and forecasted wind power by ANFIS based on 60-min average data of the last 100 samples of the test week in May

4.9 Summary

Two wind power prediction models, direct and indirect, are compared in this chapter. Since the predicted wind speed is required to forecast wind power indirectly, several time series methods based on 5-min interval were applied. DES predicted the wind speed more accurately than other statistical methods. In the next step, the proposed power curve in Chapter 3 (MHTan) aimed to determine the corresponding wind power. To set up a direct prediction of power, ANFIS and five data mining models were applied to constitute the direct wind power forecasting. In this research, a new FS method is developed by combining the mutual information and neural network to obtain only the most informative input data. Case studies on real wind park confirmed that ANFIS outperformed others. The simulation results also confirmed that direct prediction methods overall give greater performance than indirect prediction methods.

.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS FOR THE FURTHER WORK

5.1 Conclusions

High penetration of wind power into the electricity grid has called a demand for more accurate wind power prediction for the operation of wind farms. Wind turbine power curve depicting the relationship between wind speed and power can serve as a tool for wind power forecasting. Wind-power curve modeling, therefore, was set as the first objective of this thesis. To accomplish this objective, an enhanced parametric model, called modified hyperbolic tangent (MHTan) was proposed which not only forecasts the wind power but it also aids in monitoring the wind turbine to disclose any deviation from the normal performance. The monitoring feature of MHTan can significantly reduce the maintenance cost of wind turbines particularly in offshore wind farms, in which the accessibility issue is always posed. To attain the second objective of this thesis, two different methods were applied to obtain the free parameters of the MHTan. In the first method, three optimization algorithms, namely, backtracking search algorithm, cuckoo search, and particle swarm optimization were employed to obtain the best coefficients of MHTan in nine-dimensional search space through minimizing the sum of squared residuals. The other method covering the second and the third objectives was based on the distribution of the wind speed and the maximum likelihood estimation (MLE). In this context, to fulfill the third objective, five parametric and one nonparametric models were used to estimate the parameters of wind speed distribution. After close observations, it was proved that the collected wind speed followed Weibull distribution than other distributions. To test the presented models two data set were considered, one based on the actual data collected and the other based on the simulated data varied in size of samples. According to the obtained results, the nonparametric model, group search optimization, illustrated better accuracy in estimation of Weibull parameters. In the next step, a new formula was derived in which matching the frequency distribution of the turbine power. The formula comprises the known parameters of Weibull distribution and the unknown parameters of MHTan. MLE then was developed to estimate the coefficients of the MHTan. The comparative results clearly indicated that the power curve based on LSE fits the observed wind-power curve better than the one based on MLE, however, they both have the S-curve shape. The performance of the MHTan was validated by two data sets, one based on the real data collected from a pitch-control wind turbine and the other based on generated data representing horizontal-axis wind turbine with the yaw-control, and vertical-axis wind turbines with the stall- and pitch-control. The MHTan also was compared with several parametric and nonparametric models of wind turbine power curves. The numerical results affirmed the superiority of the proposed model in a pitch-control wind turbine as compared to both parametric and nonparametric models. As it was reasonably expected in the yaw- and the stall-control wind turbine as the shape of the power curve is not purely S-curve, data mining algorithms performed better than others, however, the obtained results by MHTan using BSA and CSA were absolutely satisfactory in these turbines.

This thesis also examined the performance of two wind power forecasting approaches with two different strategies. In the first one, which is called indirect prediction method, the wind speed was forecasted at first and then MHTan performed the conversion of forecasted wind speed to wind power. Double exponential smoothing method was the best as compared to other statistical methods. To set up the second model, which is called direct prediction method, six data-driven approaches were employed. The results verified that in the second method ANFIS was the best. In the second model, before applying input data such as wind speed, wind direction, temperature, a data preprocessing was required. Thus, a new feature selection technique

(FS) was introduced to select the most informative inputs. The developed FS was composed of two stages. The first stage aimed to remove the irrelevant data by mutual information, while the second stage focused on filtering out the redundant data by the neural network. The two-stage FS technique was compared to several linear and nonlinear FS methods and the comparative results proved that the developed FS technique has a greater efficiency than the others.

5.2 Recommendations for the Further Work

The following tasks are recommended for the future works.

1. In indirect prediction approaches transmission of the time series data, e,g by wavelets or Kalman filter can be considered to enhance the prediction accuracy.
2. Heuristic algorithms such as BSA or CSA can be applied in the second stage of the developed FS technique to optimize neural networks' weightings in order to avoid trapping in local minima as well as to improve its performance.
3. The MHTan or data-driven approaches can be employed in real economic dispatch problems for wind power-integrated power systems.
4. The proposed control chart by MHTan which is able to detect anomalies and outliers can be used to identify the factors and prediction of faults wherein the faults can be categorized in different groups labeled with different codes representing the severity of the problem.

REFERENCES

Aghajani, A., Kazemzadeh, R., & Ebrahimi, A. (2016). A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. *Energy Conversion and Management, 121*, 232-240.

Ai, B., Yang, H., Shen, H., & Liao, X. (2003). Computer-aided design of PV/wind hybrid system. *Renewable energy, 28*(10), 1491-1512.

Ak, R., Li, Y.-F., Vitelli, V., & Zio, E. (2018). Adequacy assessment of a wind-integrated system using neural network-based interval predictions of wind power generation and load. *International Journal of Electrical Power & Energy Systems, 95*, 213-226.

Akdağ, S., Bagiorgas, H., & Mihalakakou, G. (2010). Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Applied Energy, 87*(8), 2566-2573.

Akdağ, S. A., & Dinler, A. (2009). A new method to estimate Weibull parameters for wind energy applications. *Energy Conversion and Management, 50*(7), 1761-1766.

Al-Shammari, E. T., Amirmojahedi, M., Shamshirband, S., Petković, D., Pavlović, N. T., & Bonakdari, H. (2015). Estimation of wind turbine wake effect by adaptive neuro-fuzzy approach. *Flow measurement and instrumentation, 45*, 1-6.

Alessandrini, S., Sperati, S., & Pinson, P. (2013). A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. *Applied Energy, 107*, 271-280.

AlRashidi, M., & El-Naggar, K. (2010). Long term electric load forecasting based on particle swarm optimization. *Applied Energy, 87*(1), 320-326.

Amjady, N., Keynia, F., & Zareipour, H. (2011a). Short-term wind power forecasting using ridgelet neural network. *Electric Power Systems Research, 81*(12), 2099-2107.

Amjady, N., Keynia, F., & Zareipour, H. (2011b). Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. *IEEE Transactions on Sustainable Energy, 2*(3), 265-276.

Arslan, T., Bulut, Y. M., & Yavuz, A. A. (2014). Comparative study of numerical methods for determining Weibull parameters for wind energy potential. *Renewable and Sustainable Energy Reviews, 40*, 820-825.

Ashok, S. (2007). Optimised model for community-based hybrid energy system. *Renewable energy, 32*(7), 1155-1164.

Association, W. W. E. (2014). Half-year report. *Bonn, Germany*.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on energy conversion, 21*(1), 273-284.

Bludszuweit, H., Domínguez-Navarro, J. A., & Llombart, A. (2008). Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems, 23*(3), 983-991.

Brano, V. L., Orioli, A., Ciulla, G., & Culotta, S. (2011). Quality of wind speed fitting distributions for the urban area of Palermo, Italy. *Renewable energy, 36*(3), 1026-1039.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., & Santos, C. S. (2014). WRF wind simulation and wind energy production estimates forced by different reanalyses: comparison with observed data for Portugal. *Applied Energy, 117*, 116-126.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2): Duxbury Pacific Grove, CA.

Catalão, J. P. d. S., Pousinho, H. M. I., & Mendes, V. M. F. (2011). Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renewable energy, 36*(4), 1245-1251.

Celik, A. N. (2004). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable energy, 29*(4), 593-604.

Chang, T. P. (2011). Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy, 88*(1), 272-282.

Chedid, R., Akiki, H., & Rahman, S. (1998). A decision support technique for the design of hybrid solar-wind power systems. *IEEE Transactions on energy conversion, 13*(1), 76-83.

Chen, J., Zeng, G.-Q., Zhou, W., Du, W., & Lu, K.-D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Conversion and Management, 165*, 681-695.

Chen, N., Qian, Z., Nabney, I. T., & Meng, X. (2014). Wind power forecasts using Gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems, 29*(2), 656-665.

Chen, P., Pedersen, T., Bak-Jensen, B., & Chen, Z. (2010). ARIMA-based time series model of stochastic wind power generation. *IEEE Transactions on Power Systems, 25*(2), 667-676.

Civicioglu, P. (2013). Backtracking search optimization algorithm for numerical optimization problems. *Applied Mathematics and Computation, 219*(15), 8121-8144.

Colak, I., Sagiroglu, S., & Yesilbudak, M. (2012). Data mining and wind power prediction: A literature review. *Renewable energy, 46*, 241-247.

Council, G. W. E. (2016). Global wind report 2014. *Brussels, Belgium*, 9-10.

Croonenbroeck, C., & Dahl, C. M. (2014). Accurate medium-term wind power forecasting in a censored classification framework. *Energy, 73*, 221-232.

Damousis, I. G., Alexiadis, M. C., Theocharis, J. B., & Dokopoulos, P. S. (2004). A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on energy conversion, 19*(2), 352-361.

De Giorgi, M. G., Ficarella, A., & Tarantino, M. (2011). Error analysis of short term wind power prediction models. *Applied Energy, 88*(4), 1298-1311.

Diaf, S., Diaf, D., Belhamel, M., Haddadi, M., & Louche, A. (2007). A methodology for optimal sizing of autonomous hybrid PV/wind system. *Energy Policy, 35*(11), 5708-5718.

Dursun, B., & Alboyaci, B. (2011). An evaluation of wind energy characteristics for four different locations in Balikesir. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 33*(11), 1086-1103.

El-Fouly, T., El-Saadany, E., & Salama, M. (2007). Improved grey predictor rolling models for wind power prediction. *IET Generation, Transmission & Distribution, 1*(6), 928-937.

Elattar, E. E. (2014). Prediction of wind power based on evolutionary optimised local general regression neural network. *IET Generation, Transmission & Distribution, 8*(5), 916-923.

Erdem, E., & Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy, 88*(4), 1405-1414.

Fan, S., Liao, J. R., Yokoyama, R., Chen, L., & Lee, W.-J. (2009). Forecasting the wind generation using a two-stage network based on meteorological information. *IEEE Transactions on energy conversion, 24*(2), 474-482.

Flores, P., Tapia, A., & Tapia, G. (2005). Application of a control algorithm for wind speed prediction and active power generation. *Renewable energy, 30*(4), 523-536.

Focken, U., Lange, M., Mönnich, K., Waldl, H.-P., Beyer, H. G., & Luig, A. (2002). Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics, 90*(3), 231-246.

Fong, K., Yuen, S., Chow, C. K., & Leung, S. W. (2010). Energy management and design of centralized air-conditioning systems through the non-revisiting strategy for heuristic optimization methods. *Applied Energy, 87*(11), 3494-3506.

Garcia, A., Torres, J., Prieto, E., & De Francisco, A. (1998). Fitting wind speed distributions: a case study. *Solar Energy, 62*(2), 139-144.

Gill, S., Stephen, B., & Galloway, S. (2012). Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy, 3*(1), 94-101.

Giorsetto, P., & Utsurogi, K. F. (1983). Development of a new procedure for reliability modeling of wind turbine generators. *IEEE transactions on power apparatus and systems*(1), 134-143.

Han, L., Romero, C. E., & Yao, Z. (2015). Wind power forecasting based on principle component phase space reconstruction. *Renewable energy, 81*, 737-744.

He, S., Wu, Q. H., & Saunders, J. (2009). Group search optimizer: an optimization algorithm inspired by animal searching behavior. *IEEE transactions on evolutionary computation, 13*(5), 973-990.

Hocaoğlu, F. O., Gerek, Ö. N., & Kurban, M. (2009). A novel hybrid (wind–photovoltaic) system sizing procedure. *Solar Energy, 83*(11), 2019-2028.

Hong, Y.-Y., Chang, H.-L., & Chiu, C.-S. (2010). Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. *Energy, 35*(9), 3870-3876.

Hoolohan, V., Tomlin, A. S., & Cockerill, T. (2018). Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. *Renewable energy, 126*, 1043-1054.

Huang, Z., & Chalabi, Z. (1995). Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics, 56*(2-3), 311-322.

Ikeda, S., & Ooka, R. (2015). Metaheuristic optimization methods for a comprehensive operating schedule of battery, thermal energy storage, and heat source in a building energy system. *Applied Energy, 151*, 192-205.

Jafarian, M., & Ranjbar, A. (2010). Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine. *Renewable energy, 35*(9), 2008-2014.

Jamil, M., Parsa, S., & Majidi, M. (1995). Wind power statistics and an evaluation of wind energy density. *Renewable energy, 6*(5-6), 623-628.

Jiang, H., Wang, J., Wu, J., & Geng, W. (2016). Comparison of numerical methods and metaheuristic optimization algorithms for estimating parameters for wind energy potential assessment in low wind regions. *Renewable and Sustainable Energy Reviews*.

Jimenez, B., Durante, F., Lange, B., Kreutzer, T., & Tambke, J. (2007). Offshore wind resource assessment with WAsP and MM5: comparative study for the German Bight. *Wind Energy, 10*(2), 121-134.

Jursa, R., & Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. *International Journal of Forecasting, 24*(4), 694-709.

Justus, C., Hargraves, W., Mikhail, A., & Graber, D. (1978). Methods for estimating wind speed frequency distributions. *Journal of applied meteorology, 17*(3), 350-353.

Karaki, S., Chedid, R., & Ramadan, R. (1999). Probabilistic performance assessment of autonomous solar-wind energy conversion systems. *IEEE Transactions on energy conversion, 14*(3), 766-772.

Katinas, V., Gecevicius, G., & Marciukaitis, M. (2018). An investigation of wind power density distribution at location with low and high wind speeds using statistical model. *Applied Energy, 218*, 442-451.

Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. *Renewable energy, 34*(5), 1388-1393.

Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization.* Paper presented at the Neural Networks, 1995. Proceedings., IEEE International Conference on.

Kolhe, M., Agbossou, K., Hamelin, J., & Bose, T. (2003). Analytical model for predicting the performance of photovoltaic array coupled with a wind turbine in a stand-alone renewable energy system based on hydrogen. *Renewable energy, 28*(5), 727-742.

Kong, X., Liu, X., Shi, R., & Lee, K. Y. (2015). Wind speed prediction using reduced support vector machines with feature selection. *Neurocomputing, 169*, 449-456.

Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services, 27*, 11-23.

Kusiak, A., & Zhang, Z. (2010). Short-horizon prediction of wind power: A data-driven approach. *IEEE Transactions on energy conversion, 25*(4), 1112-1122.

Kusiak, A., Zheng, H., & Song, Z. (2009a). Models for monitoring wind farm power. *Renewable energy, 34*(3), 583-590.

Kusiak, A., Zheng, H., & Song, Z. (2009b). On-line monitoring of power curves. *Renewable energy, 34*(6), 1487-1493.

Kusiak, A., Zheng, H., & Song, Z. (2009c). Short-term prediction of wind farm power: A data mining approach. *IEEE Transactions on energy conversion*, *24*(1), 125-136.

Lahouar, A., & Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management, 103*, 1040-1051.

Landberg, L. (1999). Short-term prediction of the power production from wind farms. *Journal of Wind Engineering and Industrial Aerodynamics, 80*(1), 207-220.

Lazić, L., Pejanović, G., & Živković, M. (2010). Wind forecasts for wind power generation using the Eta model. *Renewable energy, 35*(6), 1236-1243.

Lee, Y.-S., & Tong, L.-I. (2011). Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems, 24*(1), 66-72.

Leng, H., Li, X., Zhu, J., Tang, H., Zhang, Z., & Ghadimi, N. (2018). A new wind power prediction method based on ridgelet transforms, hybrid feature selection and closed-loop forecasting. *Advanced Engineering Informatics*, *36*, 20-30.

Leung, D. Y., & Yang, Y. (2012). Wind energy development and its environmental impact: A review. *Renewable and Sustainable Energy Reviews, 16*(1), 1031-1039.

Li, S., Wang, P., & Goel, L. (2015). Wind power forecasting using neural network ensembles with feature selection. *IEEE Transactions on Sustainable Energy, 6*(4), 1447-1456.

Li, S., Wunsch, D. C., O'Hair, E. A., & Giesselmann, M. G. (2001). Using neural networks to estimate wind turbine power generation. *IEEE Transactions on energy conversion, 16*(3), 276-282.

Liu, H., Shi, J., & Erdem, E. (2010). Prediction of wind speed time series using modified Taylor Kriging method. *Energy, 35*(12), 4870-4879.

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., & Kallos, G. (2008). Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics, 96*(12), 2348-2362.

Lu, L., Yang, H., & Burnett, J. (2002). Investigation on wind power potential on Hong Kong islands—an analysis of wind power and wind turbine characteristics. *Renewable energy, 27*(1), 1-12.

Lun, I. Y., & Lam, J. C. (2000). A study of Weibull parameters using long-term wind observations. *Renewable energy, 20*(2), 145-153.

Lydia, M., Selvakumar, A. I., Kumar, S. S., & Kumar, G. E. P. (2013). Advanced algorithms for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy, 4*(3), 827-835.

Mabel, M. C., & Fernandez, E. (2008). Analysis of wind power generation and prediction using ANN: A case study. *Renewable energy, 33*(5), 986-992.

Mahmoud, T., Dong, Z. Y., & Ma, J. (2018). An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine. *Renewable energy, 126*, 254-269.

Maldonado, S., Carrizosa, E., & Weber, R. (2015). Kernel Penalized K-means: A feature selection method based on Kernel K-means. *Information Sciences, 322*, 150-160.

Marvuglia, A., & Messineo, A. (2012). Monitoring of wind farms' power curves using machine learning techniques. *Applied Energy, 98*, 574-583.

Meng, X. Modeling and Forecasting Hourly Wind Power Production in Sweden with Time Series Models.

Mohandes, M., Rehman, S., & Rahman, S. (2011). Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS). *Applied Energy, 88*(11), 4024-4032.

Mohandes, M. A., Halawani, T. O., Rehman, S., & Hussain, A. A. (2004). Support vector machines for wind speed prediction. *Renewable energy, 29*(6), 939-947.

Montgomery, D. C. (2009). *Introduction to statistical quality control*: John Wiley & Sons (New York).

Nelson, D., Nehrir, M., & Wang, C. (2006). Unit sizing and cost analysis of stand-alone hybrid wind/PV/fuel cell power generation systems. *Renewable energy, 31*(10), 1641-1656.

Nguyen, T. T., Vo, D. N., & Truong, A. V. (2014). Cuckoo search algorithm for short-term hydrothermal scheduling. *Applied Energy, 132*, 276-287.

Nielsen, H. A., Nielsen, T. S., Madsen, H., Pindado, M. J., & Marti, I. (2007). Optimal combination of wind power forecasts. *Wind Energy, 10*(5), 471-482.

Niu, D., Wang, Y., & Wu, D. D. (2010). Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications, 37*(3), 2531-2539.

Nourani Esfetang, N., & Kazemzadeh, R. (2018). A novel hybrid technique for prediction of electric power generation in wind farms based on WIPSO, neural network and wavelet transform. *Energy, 149*, 662-674.

Ohunakin, O., Adaramola, M. S., & Oyewola, O. M. (2011). Wind energy evaluation for electricity generation using WECS in seven selected locations in Nigeria. *Applied Energy, 88*(9), 3197-3206.

Palomares-Salas, J., De La Rosa, J., Ramiro, J., Melgar, J., Aguera, A., & Moreno, A. (2009). *ARIMA vs. Neural networks for wind speed forecasting.* Paper presented at the Computational Intelligence for Measurement Systems and Applications, 2009. CIMSA'09. IEEE International Conference on.

Palomares-Salas, J., De la Rosa, J., Ramiro, J., Melgar, J., Agüera, A., & Moreno, A. (2009). *Comparison of models for wind speed forecasting.* Paper presented at the Proc. International Conference on Computational Science.

Park, J.-Y., Lee, J.-K., Oh, K.-Y., & Lee, J.-S. (2014). Development of a novel power curve monitoring method for wind turbines and its field tests. *IEEE Transactions on energy conversion, 29*(1), 119-128.

Pearre, N. S., & Swan, L. G. (2018). Statistical approach for improved wind speed forecasting for wind power production. *Sustainable Energy Technologies and Assessments, 27*, 180-191.

Peng, H., Liu, F., & Yang, X. (2013). A hybrid strategy of short term wind power prediction. *Renewable energy, 50*, 590-595.

Piechocki, J., Ambroziak, D., Palkowski, A., & Redlarski, G. (2014). Use of Modified Cuckoo Search algorithm in the design process of integrated power systems for modern and energy self-sufficient farms. *Applied Energy, 114*, 901-908.

Pinson, P. (2006). *Estimation of the uncertainty in wind power forecasting.* École Nationale Supérieure des Mines de Paris.

Pousinho, H. M. I., Mendes, V. M. F., & Catalão, J. P. d. S. (2011). A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal. *Energy Conversion and Management, 52*(1), 397-402.

Powell, W. R. (1981). An analytical expression for the average output power of a wind machine. *Solar Energy, 26*(1), 77-80.

Raichle, B. W., & Carson, W. R. (2009). Wind resource assessment of the southern appalachian ridges in the southeastern united states. *Renewable and Sustainable Energy Reviews, 13*(5), 1104-1110.

Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., Monteiro, C., Sousa, J., & Bessa, R. (2009). Comparison of two new short-term wind-power forecasting systems. *Renewable energy, 34*(7), 1848-1854.

Rocha, P. A. C., de Sousa, R. C., de Andrade, C. F., & da Silva, M. E. V. (2012). Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. *Applied Energy, 89*(1), 395-400.

Saleh, A. E., Moustafa, M. S., Abo-Al-Ez, K. M., & Abdullah, A. A. (2016). A hybrid neuro-fuzzy power prediction system for wind energy generation. *International Journal of Electrical Power & Energy Systems, 74*, 384-395.

Schlechtingen, M., Santos, I. F., & Achiche, S. (2013). Using data-mining approaches for wind turbine power curve monitoring: a comparative study. *IEEE Transactions on Sustainable Energy, 4*(3), 671-679.

Senjyu, T., Yona, A., Urasaki, N., & Funabashi, T. (2006). *Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator.* Paper presented at the Power Systems Conference and Exposition, 2006. PSCE'06. 2006 IEEE PES.

Setiono, R., & Liu, H. (1997). Neural-network feature selector. *IEEE transactions on neural networks, 8*(3), 654-662.

Sfetsos, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series. *Renewable energy, 27*(2), 163-174.

Sharifian, A., Ghadi, M. J., Ghavidel, S., Li, L., & Zhang, J. (2018). A new method based on Type-2 fuzzy neural network for accurate wind power forecasting under uncertain data. *Renewable energy, 120*, 220-230.

Shi, Y., & Eberhart, R. (1998). *A modified particle swarm optimizer*. Paper presented at the Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on.

Shokrzadeh, S., Jozani, M. J., & Bibeau, E. (2014). Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy, 5*(4), 1262-1269.

Sideratos, G., & Hatziargyriou, N. (2007). *Using radial basis neural networks to estimate wind power production.* Paper presented at the Power Engineering Society General Meeting, 2007. IEEE.

Sideratos, G., & Hatziargyriou, N. D. (2007). An advanced statistical method for wind power forecasting. *IEEE Transactions on Power Systems, 22*(1), 258-265.

Sideratos, G., & Hatziargyriou, N. D. (2012). Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems, 27*(4), 1788-1796.

Sterba, J., & Hilovska, K. (2010). The implementation of hybrid ARIMA neural network prediction model for aggregate water consumption prediction. *Aplimat—Journal of Applied Mathematics, 3*(3), 123-131.

Taylor, J. W., & McSharry, P. E. (2007). Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems, 22*(4), 2213-2219.

Thanaa, F., Eskander, M. N., & El-Hagry, M. T. (2006). Energy flow and management of a hybrid wind/PV/fuel cell generation system. *Energy Conversion and Management, 47*(9), 1264-1280.

Thapar, V., Agnihotri, G., & Sethi, V. K. (2011). Critical analysis of methods for mathematical modelling of wind turbines. *Renewable energy, 36*(11), 3166-3177.

Torres, J. L., Garcia, A., De Blas, M., & De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy, 79*(1), 65-77.

Tuev VI, U. M. (2009). Using modified hyperbolic tangent function to approximate the current-voltage characteristics of field-effect transistors. *Tomsk Politehnic University, p. 135–8*.

Ucar, A., & Balo, F. (2009). Evaluation of wind energy potential and electricity generation at six locations in Turkey. *Applied Energy, 86*(10), 1864-1872.

Usta, I. (2016). An innovative estimation method regarding Weibull parameters for wind energy applications. *Energy, 106*, 301-314.

Velo, R., López, P., & Maseda, F. (2014). Wind speed estimation using multilayer perceptron. *Energy Conversion and Management, 81*, 1-9.

Xu, Q., He, D., Zhang, N., Kang, C., Xia, Q., Bai, J., & Huang, J. (2015). A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on Sustainable Energy, 6*(4), 1283-1291.

Yang, H., Lu, L., & Burnett, J. (2003). Weather data and probability analysis of hybrid photovoltaic–wind power generation systems in Hong Kong. *Renewable energy, 28*(11), 1813-1824.

Yang, H., Wei, Z., & Chengzhi, L. (2009). Optimal design and techno-economic analysis of a hybrid solar–wind power generation system. *Applied Energy, 86*(2), 163-169.

Yang, L., He, M., Zhang, J., & Vittal, V. (2015). Support-vector-machine-enhanced markov model for short-term wind power forecast. *IEEE Transactions on Sustainable Energy, 6*(3), 791-799.

Yang, X.-S., & Deb, S. (2009). *Cuckoo search via Lévy flights.* Paper presented at the Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on.

Yesilbudak, M., Sagiroglu, S., & Colak, I. (2013). A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Conversion and Management*, *69*, 77-86.

Zeng, J., & Qiao, W. (2012). Short-term wind power prediction using a wavelet support vector machine. *IEEE Transactions on Sustainable Energy, 3*(2), 255-264.

Zhang, Z. (2012). *Performance optimization of wind turbines*: The University of Iowa.

Zhou, J., Erdem, E., Li, G., & Shi, J. (2010). Comprehensive evaluation of wind speed distribution models: A case study for North Dakota sites. *Energy Conversion and Management, 51*(7), 1449-1458.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

JOURNAL PAPERS

Renani, E. T., Elias, M. F. M., & Rahim, N. A. (2016). Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management*, *118*, 193-203 (ISI, Q1)

Taslimi-Renani, E., Modiri-Delshad, M., Elias, M. F. M., & Rahim, N. A. (2016). Development of an enhanced parametric model for wind turbine power curve. *Applied Energy*, *177*, 544-552 (ISI, Q1)

CONFERENCE PAPER

Taslimi-Renani, E., Modiri-Delshad, M., Elias, M. F. M., & Rahim, N. A., "Wind Power Prediction Using Enhanced Parametric Wind Power Curve Modeling," *in Clean Energy and Technology (CEAT)*, 2016 IET Conference on,